NASSDOC/GIDS TRAINING PROGRAMME

on

INFORMATION RESOURCE MANAGEMENT FOR SOCIAL SCIENTISTS IN EMERGING KNOWLEDGE ERA
13-15 FEBRUARY, 2007

PROGRAMME SCHEDULE

ADDITIONAL READING MATERIAL

LIST OF RESOURCE PERSONS

LIST OF PARTICIPANTS

GIRI INSTITUTE OF DEVELOPMENT STUDIES
LUCKNOW

ICSSR/NASSDOC
Training Programme
on
INFORMATION REOURCE MANAGEMENT FOR
SOCIAL SCIENTISTS IN EMERGING KNOWLEDGE ERA
13-15 FEBRUARY 2007
PROGRAMME SCHEDULE
Registration-09.30-10.30 A.M.

13.02.2007 (Tuesday)

Inaugural Session	Session I		Session II		Session III		Session IV
10.30 A.M.-11.15. A.M. Chief Guest Prof. Prem Vrat, Vice Chancellor, U.P. Technical University	11.30 A.M.-12.30 P.M. An Overview of Social Science Web Resources Dr. A Sood, JNU, New Delhi	Tea: 12.30-12.45 P.M.	12.45 P.M-01.45 P.M. Internet and Social Science Research: ISID Experience Prof. K.V.K. Ranganathan, ISID, New Delhi	Lunch 01.45-02.30 P.M.	02.30 P.M.-03.30 P.M. Practical Session Ms. Rita Sinha Ms.A. Srivastava	Tea: 03.30-03.45 P.M.	03.45 P.M.-05.00 P.M. Digital Libraries: An Overview Dr. N.R. Satyanarayan

14.02.2007 (Wednesday)

Session I		Session II		Session III		Session IV
10.00 A.M.-11.30 A.M. Techniques and Search Strategies of Subject Gateway and Portals Dr. Rochana Srivastava	Tea 11.30-11.45	11.45 A.M.-01.15 P.M. Management of Web Resources in Social Sciences Dr. R. Chaddha	Lunch 01.15-02.00	02.00 P.M.-03.30 P.M. Building Digital Library Collection for Social Scientists Dr. R. Chaddha	Tea 03.30-03.45	03.45 P.M.-05.00 P.M. Practical Session Ms. Rita Sinha Ms. Lata Bajpai

15.02.2007 (Thursday)

Session I		Session II		Session III		Session IV	Session V
10.00 A.M.11.30 A.M. Information Resources: A case study of GIDS library Dr. Anuradha Kakkar	Tea 11.30-11.45	11.45 A.M.-01.15 P.M. Information Resources for Official Publications Dr. Rakesh Srivastava	Lunch 01.15-02.00	02.00 P.M.-03.30 P.M. Identification and Access to e-resources Dr. Rakesh Srivastava	Tea 03.30-03.45	03.45 P.M.-04.00 P.M. Feed Back	04.00 P.M.-05.30 P.M. Valediction Prof. S. Chakraborty, Director, Jaipuria Instt. of Management

Resource Faculty

- Prof. Atul Sood-Prof., JNU, New Delhi.

- Dr. Ravindra Chaddha-Jt. Sec., Lok Sabha Secretariat, Parliament Library., New Delhi-1.

- Dr. Rakesh Srivastava, Chief Librarian, Supreme Court, New Delhi.

- Dr. Rochana Srivastava, HOD, Dept. of Library & Information Science, I.T. College, Lucknow.

- Prof. K.V.K. Ranganathan-Institute of Studies in Industrial Development, New Delhi.

- Dr. N.R. Satyanarayana-Prof. & Head, Dept. of Library & Information Science, L.U. Lucknow.

- Programme Director-Prof. A. Joshi

- Course Coordinato-Dr. Anuradha Kakkar

Practical Session

- Ms. Rita Sinha

- Ms. Anjali Srivastava

NASSDOC/GIDS TRAINING PROGRAMME

on

INFORMATION RESOURCE MANAGEMENT FOR SOCIAL SCIENTISTS IN EMERGING KNOWLEDGE ERA

BACKGROUND MATERIAL

Social Science Information Services:
Some Reflections

Harsha Parekh
Visiting Professor
SHPT School of Library Science
SNDT Women's University
Mumbai 400 020

The opportunity of combining teaching and research with practical experiences of running a university library and an information centre with special emphasis on the social sciences has resulted in certain perceptions relating to information provision services to and in the social sciences. It is this implicit knowledge that I have gained over the years that I propose to share in this paper/presentation.

Information provision is one aspect of the scholarly communication process. This communication process is one through which scholars convey their knowledge to, and exchange ideas with, each other and future generations; it refers to a "conceptualisation of the process whereby a valuable raw material, namely ideas, turns into a digestible consumer product, information and the higher form of information we call knowledge."[1]

A discipline represents a community of scholars who communicate with each other to enrich their thought processes, to share hunches, observations and perceptions as well as to ensure critique and objectivity. The process is an interactive one spanning geographical boundaries and crossing time periods. It sets out not only the very boundaries of a discipline, but establishes the common vocabulary and paradigms within which that community or profession functions.

As a preliminary to understanding the issues, I first begin with some characteristics of the field of social science and then move on to perceptions of social scientists' need for information and their information seeking behaviour patterns. Finally I conclude with suggestions for service provision.

The Social Sciences

Social sciences, refers to a conglomerate of subjects; some of these are pure disciplines (economics) while others are interdisciplinary (women's studies); some are more theoretical (sociology) while others are focussed on applications.

Although some generalisations are possible, the individual subject fields vary in many ways. By their very definition, these sciences are "social" and deal with issues in society. Many of these concepts are limited to certain cultures and terms used to represent them are not universal. In matters concerning concepts and terminology the difference between the natural and social sciences is very sharp. In the pure and applied natural sciences, concepts are clear. This means that when communicating about them, a great deal of commonality and sharing can be achieved. Also, when the

concept alters, new terms are frequently created to refer to them; the scientific community is ready to accept useful neologisms. The result is that terms retain meaning and do not become "rubbery".

In the social sciences and humanities, more concepts are culture dependent. Concepts such as guru, sati, devdasi are restricted to Indian culture; such restricted local concepts are found in many disciplines e.g. "jajmani" in economics, "dharma" in religion, "gramsevika" and "panchayat" in local government. Western born terms are inadequate and new terms are coined to represent them; a good example is the word "sanskritization" in sociology.

In the social sciences, there seems to be an aversion to neologisms. Older terms are used to mean something different e.g. "streak" or "gay" or even "exit policy". Since terms are rubbery, when researchers or theorists require a more precise meaning, they assign or stipulate a specific meaning. (Thus in almost all social science research one finds a section or chapter on "definitions"; this is often not required in the natural sciences.). Further, subject vocabularies are often not agreed upon and a term is sometimes used in different contexts e.g. 'community'.

Concepts are linked together into theories; a subject is concerned with numerous theories. Frequently, one or several theories are viewed within a certain frame or paradigm. Kuhn, who elaborated on the paradigms of science, believed that in normal times, the work of a scientist, is concerned with filling out details within the frame; occasionally, a theory is put forth which upsets the paradigm[2]. Such a scientific revolution calls forth a completely new frame. Examples of such scientific revolutions are provided by Copernicus, Wegener and his theory of continental shift, Einstein, etc.

In contrast to the pure sciences, where although several frames may be in use at the same time, there is a high degree of consensus about theories and paradigms, in the social sciences, there is a very low level of consensus about theories and methods. Plurality of paradigms gives rise to several alternate schools of thought which co-exist.

Social Science Information (SSI)
Social Science Information (SSI) is neither a clear nor a cohesive concept. It can mean either information on social science or that needed by social scientists. In either case there is no clarity; many others than social scientists – including policy makers, media personnel and the community at large, use information on social sciences. On the other hand, social scientists use a lot of information, which technically rests outside the realm of social science.

Whichever way we interpret the term – SSI – there are differences between the individual subject areas. The scholarly communication process within the fields differs in many ways e.g. psychology has more journals than political science, economics depends more on its own literature, while social anthropology cites considerably from other social science disciplines. A growing field like women's studies has more newsletters and conference literature than established fields that

have developed further specialisations, each with its range of journals. The number of national and international conferences in India on women related issues during the period 1975 to 1985 (10 a year) rose three times in the next decade (30 a year).

Although there are finer differences, some common patterns are generally visible across subject areas. Since there is emphasis on local matters and the immediate context, the use of international literature is counterbalanced with the need for local concepts, data and literature. To summarize, the characteristics of the primary literature of social sciences are:

- of varied formats
- frequently not published through formal, commercial channels
- frequently project specific
- difficult to obtain e.g. activist literature, commissioned reports, etc.

Each discipline, as it is established, develops its own secondary literature. This literature consists of substantive compilations for ready reference and information and a bibliographic apparatus to locate material. The range of secondary literature is depicted in Figure 1.

Figure 1: Range of Secondary Literature

Substantive				Indicative
1	2	3	4	5
Dictionaries	News Summaries	Trend Reports	Reviews	Indexes
Yearbooks	Encyclopedias	Bibliographic	Abstracts	Bibliographies
Statistical	Biographical	Essays	Annotated	Catalogues
Yearbooks	Dictionaries	Directories of	Bibliographies	Union Catalogues
		Completed	Guides to	Current Awareness
		Research	Literature	Service
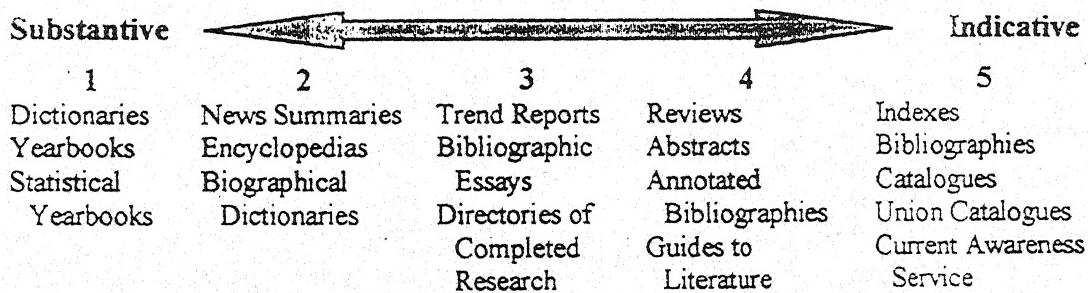
In the social sciences, with a few exceptions, comprehensive secondary sources for Indian literature are scarce. Most of the secondary tools are developed by individual organisations and based on individual collections.

Users and their Use Behaviour
There are several user groups who need information in the social sciences. At least five major groups who need information can be identified.

- **Research, Education & Training Institutes**
  1. Research Organizations
  2. Universities
     a. Teachers
     b. Students
     c. Researchers
     d. Curriculum designers

3　Colleges
　　4　Other Education Institutions
　　5　Training Institutes
◆　**Government (at National, State & Local Levels)**
　　1　Legislators/Politicians
　　2　Bureaucrats
◆　Government Bodies/organization Media
　　1　Newspapers
　　2　Magazines
　　3　Radio
　　4　:Television
◆　**Other Organizations**
　　1　Voluntary organizations
　　2　Corporations
　　3　Social action groups
　　4　Advocacy groups
　　5　Mahila Mandals
◆　**Individuals**
　　1　Professionals
　　2　Ordinary citizens

Non-Academic Users

The behaviour of many of these groups differ considerable. Most groups – other than the academic groups – very frequently are unable to recognise their information needs. For the individual, the need for information per se is not something distinct like the need for food or water, which is recognized. Information needs arise out of other basic physiological, affective and cognitive needs. Very few individuals go out and specifically look for information in libraries.

The nature of events, social factors such as socio-economic status and homogeneity of the community, cultural values, and behavioural styles affect information receptivity and search. For example, Vishwanath et al report[3] that in case of startling events (accidents, test matches, etc) there is a more uniform receiving of information, although there may be differences in in-depth knowledge. Miller categorizes people according to their style of responding to periods of acute stress. Some individuals, like to be in command of the situation, to monitor it and these people actively search out information. Others wish to distance themselves from the stress and blunt the pain and do so by not deliberately seeking information.[4]

Finally, it is necessary to recognize that need for information is frequently accompanied by the need for some element of guidance or advice. Very often it is difficult to distinguish between the two. Distinction has been made between the need for information to reduce "uncertainty" (the absence of information) and "equivocality" (ambiguity of the information). Library resources such as formal documents, numerical data and the like, are classified as "lean" and are best for reducing uncertainty. On the other hand, "rich media" such as meetings and personal face-to-face contact are necessary to reduce equivocality, because they facilitate

debate and the clarification of issues, and enable individuals to enact a common workable perception of their problem.[5]

Academic Users

Generally, the academic and research groups, for whom searching for information is part of the teaching-learning-research process prefer to undertake their own searches rather than delegating the work to assistants. There are several reasons for this. Firstly, this is because the terms have different shades of meaning. Secondly, the approach, theoretical orientation and purpose of the literature are as important as the 'subject'. A book analysing a bank strike, from the perspective of decision-making theories it exemplifies would not be very useful to a student of labour economics. A researcher studying the problem from a free trade perspective would not, in most probability, use a paper on international economics from a Marxist orientation. Similarly, researchers, activists, voluntary organisations would use different documents on, say, dowry. Since search tools are based on science models, they depend on keywords that reflect the content rather than the approach or orientation. Satisfaction with using the traditional bibliographical tools is low among the social scientists, resulting in a comparatively low use.

In the actual search behaviour, two factors influence the activity. People think within a certain template. We learn through a process of template 'adjustment', we begin with a stable template, our information need creates a situation in which are willing to change the template, based on the information we do so; the template freezes again till a new 'need' is created. Thus the cognitive states of our mind, alternately vary between stable and variable. Further our needs could be focussed or ill defined.

Viewed thus it is possible to distinguish between four different kinds of information seeking behaviour.

Figure 2: Information Needs

Cognitive State	Well defined concept	Ill defined concept
Stable Cognitive State	1 Limited uncertainty Relevance Assessment : Yes Curiosity Low Confined Navigation	4 High uncertainty Relevance Assessment : No Curiosity Low Dead end
Variable Cognitive State	2 Controlled uncertainty Relevance Assessment : Yes Curiosity High Exploratory Navigation	3 High uncertainty Relevance Assessment : No Curiosity High Browsing

Ingwersen and Willett[6] distinguish between the information behaviour as they occur in the situations identified in the above boxes 1, 2, and 3. In the first case the individual needs to verify or confirm what (s)he already knows or understands. Therefore there is very little desire to extend or broaden the search; Ingwersen and

Willets refer to these needs as "verificative needs". How often have librarians come across readers who do not wish to go beyond what they came looking for.

In the second case, the user's variable cognitive state indicates that (s)he is willing to change his frame of reference or mental schema about a topic; these needs are "conscious topical needs". We have experienced examples of both these types: in one case the user comes looking for "wife-battering" and is unwilling even to look at other information on domestic violence; a second user searching for information on the same topic "sees" a connection between the topic and alcoholism and the topic of the need is widened. Similarly a researcher looking for information about women in the informal sectors stretches her need to women entrepreneurs.

Case 3 is an inability to create a focus to the need; one of "Muddled Needs". It results in a vaguely expressed demand, for example I need a bibliography on "Women's problems" or "women's issues". Through browsing or question negotiation, however, a focus can be reached. Box 4 is a dead end, which does not even recognize the need.

Planning Information Services

All information services have two fundamental objectives: (i) helping users identify useful documents thereby improving the flow of bibliographic information and (ii) improve access to documents.

Today, the context in which these activities have to be conducted has undergone revolutionary changes under the influence of information technology. All aspects of computer technology - processing techniques, storage media and input/output devices - have become faster, more compact, sophisticated and less expensive. Accompanying these developments are the innovations in telecommunications -- in transmission media, switching technology and terminal equipment. The convergence of different technologies is creating new contexts, new opportunities and new needs.

The rapid emergence and development of electronic information technologies make it possible to envision radically different ways or organising collections and services traditionally provided by libraries. The crux of the difference lies in the nature of the digital document and the possibility of remote access.

As the Mellon report points out that "We have lived for many generations with a world in which the technology of publication meant that access required ownership, in other words, that scholarly communication was usable only if it were gathered in a large site-specific, self-sufficient collection... New electronic technologies allow the possibility of uncoupling ownership from access, the material object from its intellectual content. This possibility is revolutionary, perhaps dramatically so."[8]

The change has been described as equivalent to a Copernican revolution — the library is no longer the centre of the information provision world; there are many providers and it is the user who has become the centre. (Figure 3)

Figure 3: The Copernican Revolution in Information Provision
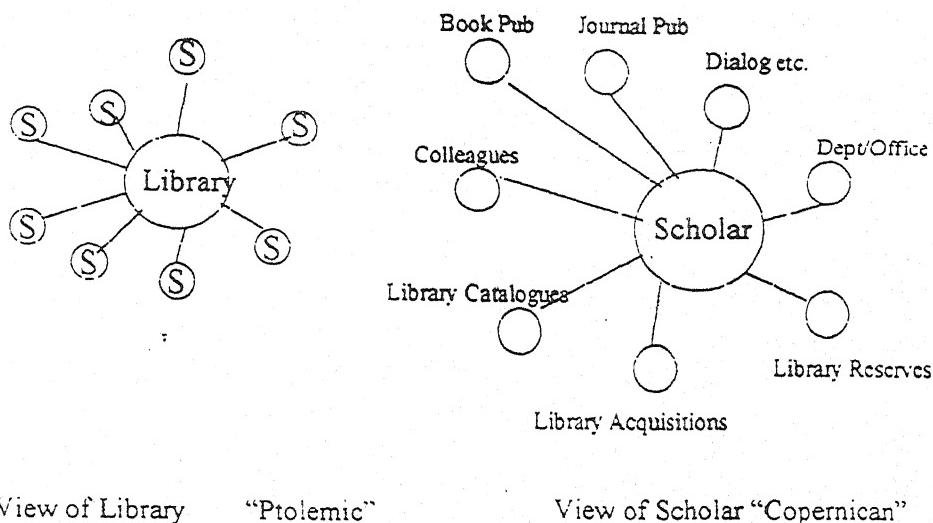
View of Library   "Ptolemic"   View of Scholar "Copernican"

A user-oriented approach requires that the context be kept in mind and a good need-service match takes place. Based on some of the points made above, I would like to make a few suggestions.

Coverage

Since social scientists use a mix of international and local publications emphasis should be on covering both these satisfactorily. The growth in the number and formats of publications, the increase in specialisations and the rise in prices make acquiring international information materials a challenge. To meet the challenge, careful planning, sharing of responsibilities, networking, consortia approc?? have to be worked out at a national level. My impressions are that librari?? ?ai are better equipped (perhaps, with the exception of econor?? ??s published in the 1970s than 1990s.

Local materials, including that in local langua???? have to be collected. A systematic effort to collect elusive materials has to be made, since these are sources of future research. If for example, twenty years from date, if a political scientist wishes to do research on the 2004 general elections, (s)he will look for party manifestos, posters, campaigning materials in different languages. To satisfy the future scholar, steps need to be taken today.

I would like to mention two other important categories of sources. Digital information sources are growing in importance and the need to create reservoirs of such information is fast emerging. The technology is also providing opportunities to build open archives, which need to be seized. As indicated earlier, bibliographic tools in India are limited often to institutional initiatives. Attempts to collect them at the national level have to be initiated.

Organisation of Information

Three issues here need to be highlighted. Firstly, when providing access points several additional tags other than the primary keywords would certainly help. This is particularly true of research which spans two or more disciplines. For example if a study on primary health care needs of tribal women were to be indexed and included in local indexes, the needs of scholars from the fields of social anthropology, women's studies and sociology should be kept in mend. Similarly, the need for tags for target audience, the context, ideology, age group of persons dealt with (as in psychology) etc. could be considered.

Secondly, since social science literature covers many local concepts, the indexing tools need to be modified to include terms for many of these. There is an urgent need to develop a list of subject headings and/or a thesaurus to establish vocabulary control for local concepts. The need to develop multi-lingual bibliographic control tools is another related area which needs to be attended to.

A third issue arises when we consider the development of resource discovery tools on the Internet. The reorganization of the full text electronic documents using a database management system increases access routes to them. This requires that each item be provided with its own metadata thereby improving their retrievability. Further, a digital library needs to be was organised both for browsing through its broad groupings and hyperlinks and for searching through a search engine.

Services

Several international studies have indicated that social scientists are involved with dissemination activities relating to a particular research project for a longer duration; however, their interest and focus changes more often than the pure scientist's does. The social scientist uses less of selective dissemination services, but undertakes more retrospective searches. Readers' services – help in locating information, guidance, translation, are equally necessary to the social scientist as to the pure scientist. Perhaps a little more help in articulating their needs and search formulations would also be required.

The opportunity NASSDOC has of considering afresh, the sources and services required cannot be wasted. Interaction with users is essential at this stage; so also is looking at the individual disciplines and the finer differences that exist between them.

[1] Duff, Alistar: "Models of Communication" *Journal of Library and Information Science* v.29 (4) December 1997. p. 179.

[2] Kuhn, T S: The Structure of Scientific Revolutions. 2nd ed. Chicago: University of Chicago Press. 1973.

[3] Vishwanath, K et al : Health and knowledge gaps : some lessons from the Minnesota Heart Health Program. American Behavioral Scientist 34 (6) July-August 1991. p.712-726.

[4] Miller, Suzanne M : Monitoring and blunting : validation of a questionnaire to assess styles of information seeking under threat. Journal of Personality and Social Psychology v.52 (1987) p.345-353. Cited by Barker, Lynda : A Study of the nature of information needed by women with multiple sclerosis. Library and Information Science Research v.18 (1) Winter 1996. p.67-81.

[5] Daft, R L and Lengel, R H : Organizational information requirements, media richness and structural design. Management Science 1986. 32. p 554-571. Cited by Kaye, David: An Information Model of Organization. Managing Information

[6] Ingwersen, Peter & Willett, Peter : An Introduction to Algorithmic and cognitive approaches to Information Retrieval. Libri 45 (3/4) Sept/Dec 1995. p.160-177.

[8] University Libraries and Scholarly Communication: A Study prepared for the Andrew W Mellon Foundation, 1992. Reprinted in Journal of Library Administration v.23 (3/4) 1996 p. xxiv.

NEW DATABASE PRODUCTS: SOCIAL SCIENCE, HUMANITIES, NEWS AND GENERAL (ISSUE 14)

NEW DATABASE PRODUCTS

Martha E. Williams

Professor of Information Science, Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email: m-will-13@uiuc.edu

Daniel E. Burgard

Education and Social Science Library, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
Email: dburgard@uiuc.edu

Introduction

This is the fourteenth article on social science, humanities, news and general databases in a continuing series of articles summarizing and commenting on new database products. Two companion articles, one covering science, technology and medicine (STM) appeared in Online & CD-ROM Review vol. 23, no. 4 and the other covering business and law (BSL) will appear in Online & CD-ROM Review vol. 23, no. 6. The articles are based on the newly appearing database products in the Gale Directory of Databases.[1] The Gale Directory of Databases (GDD) was created in January 1993 by merging Computer-Readable Databases: A Directory and Data Sourcebook (CRD)[2] together with the Directory of Online Databases (DOD) and the Directory of Portable Databases (DPD).

Organization

The Gale Directory of Databases (hereinafter to be referred to as the Directory) is a two-volume work published twice each year. This article and its companions are based on the fourteenth issue of the Gale Directory of Databases, the March 1999 issue (as of September 1997 Gale Research (now named The Gale Group) changed the publication dates of the semiannual issues of the Directory to September and March rather than January and July as it had been in the past). Three 'new database products' [3] articles corresponding to each issue of the Directory and on each of three subject groupings of databases — STM, SSH and BSL — are published twice per year in Online & CD-ROM Review — one article per issue for each of the six issues of the journal that are published each year. The second set of three articles published in 1999 — one each per subject group — correspond to Issue 14 of the Directory, with the Directory issue number appearing in parentheses after the subtitle. Articles on all three subject groups [4] are published twice per year, thereby keeping readers continuously up-to-date.

Social Science, Humanities, News and General Databases

The full list of new SSH database products is given in Table 2 following the text of this article. Included for each entry are database name, vendor, medium and an indication of whether or not the database is totally new (New) or newly implemented (NI). There were 143 new SSH database products of which 70 per cent were totally new databases and 30 per cent were new publications of existing databases in the same medium but by a different vendor or in a new medium.

10

New Database Products

Table 1, 'Distribution and percentage of new and new implementation social science, humanities, news and general database products by medium', lists the various media — batch, CD-ROM (including other optical), diskette, handheld, magnetic tape and online — and indicates for each the number of new or newly implemented SSH databases that are listed in GDD Issue 14.

There were 143 'New Database Products: social science, humanities, news and general' in Issue 14 in comparison with 251 in issue 13. The percentage of totally new databases versus new implementations of existing databases decreased. In issue 14, 70 per cent of the entries are new databases and 30 per cent are new implementations (versus 80 per cent new and 20 per cent new implementations in Issue 13). There were: 1 new handheld product, 2 new batch and 12 new magnetic-tape products, 11 new diskette products, 76 new CD-ROM products and 41 new online products. Fifty-three percent of the new products were on CD-ROM (versus 56 per cent in Issue 13) and 29 per cent were online (versus 34 per cent in Issue 13). Diskettes changed from 4 per cent in issue 13 to 8 per cent in Issue 14.

Table 1. *Distribution and Percentage of New and New Implementation Social Science, Humanities, News and General Database Products by Medium*

Medium	New	(per cent)	NI*	(per cent)	Total	(per cent)
Batch	0	(0)	2	(01)	2	(01)
CD-ROM	59	(41)	17	(12)	76	(53)
Diskette	5	(04)	6	(04)	11	(08)
Handheld	0	(0)	1	(01)	1	(01)
Mag.Tape	3	(02)	9	(06)	12	(08)
Online	33	(23)	8	(06)	41	(29)
Total	100	(70)	43	(30)	143	(100)

*New Implementation

Magnetic Tape and Batch

Many well-established databases are now offered on magnetic tape. **EconLit**, the premier index to economic literature, is available on tape from OVID. EconLit is produced by the American Economic Association and offers citations and abstracts from over 500 journals beginning in 1969. EconLit corresponds to the print publications **Journal of Economic Literature**, the **Index of Economic Articles** and **Abstracts of Working Papers in Economics**. It covers all areas of economics, including theory, history, quantitative, international, public sector, inflation, trade, markets, social indicators and credit. In addition to being available from OVID, EconLit is also available from a variety of other vendors such as OCLC, EBSCO and SilverPlatter.

Another new magnetic tape offering from OVID is **sociofile**. Now owned by Cambridge Scientific Abstracts, sociofile indexes journals, conference papers and dissertations in sociology and related disciplines. Comprising *Sociological Abstracts* and *Social Planning, Policy, and Development Abstracts* (SOPODA), sociofile begins its coverage in 1963 and indexes the contents of over 2500 journals. sociofile covers all areas of sociology, from the theoretical to the applied and is available from a variety of vendors in a range of formats.

The H.W. Wilson product **Social Sciences Abstracts Full Text** is now available on magnetic tape. This database contains about 550,000 citations from over 400 periodicals covering a range of social science topics. Among the subjects covered are anthropology, economics, geography, law, public policy, psychology, sociology and social work. All included periodicals are indexed from 1983 onward while abstracts are available for articles from 1994 on and full-text from January 1995 onward. Again, **Social Sciences Abstracts Full Text** is available from numerous vendors such as SilverPlatter, OVID and Dialog.

Japanese Periodicals Index is produced and sold by the Japan National Diet Library. Its content is drawn from over 5000 Japanese academic journals and its subject coverage ranges from science and technology to the social sciences and the humanities. Users will find over 1.8 million citations to articles from 1985 to date in the **Japanese Periodicals Index**.

The **Gale Guide to Internet Databases** is an exciting product now available on magnetic tape and diskette. It provides access to over 4000 Internet-accessible databases and focuses on government, academic, research and educational topics. Pop culture sites are also included. Sites with all types of data formats can be found here, including numeric, bibliographic, directories, bulletin boards etc. For each site, URLs, ftp addresses and email addresses are given. The **Gale Guide to Internet Databases** corresponds to the print publication *Gale Guide to Internet Databases* and the online *Gale Directory of Online, Portable, and Internet Databases*.

CHARACTERISTICS OF SOCIAL SCIENCE INFORMATION: A SELECTIVE REVIEW OF THE LITERATURE. PART II

H. P. HOGEWEG-DE HAART*

Gerrit van der Veenstraat 122, 1077 E R Amsterdam, The Netherlands

ABSTRACT

Part II of this review examines studies of the users of social science information: both researchers and practitioners, from the INFROSS studies of the late 1960s and early 1970s to more recent studies of practitioners in various fields. General conclusions and implications are drawn from this review.

CHARACTERISTICS OF THE USE OF SOCIAL SCIENCE INFORMATION

Introduction

Contrary to scientific information which is used almost exclusively by scientific researchers, social science information is of interest to various categories of users: social science researchers, practitioners and policy-makers. The flow of information between basic research and practitioners as well as policy-makers is quite unsatisfactory. Brittain (1970), referring to a study of Mersel *et al.* (1966), in which the model that identifies universities with creators of research projects, and practitioners with consumers of the products is criticized, concludes:

> This leaves practitioners very ignorant about the mass of basic research and basic researchers very often ignorant (and/or indifferent) of the needs of practitioners (Brittain, 1970:123).

Discussing information needs, Wilson distinguishes between physical needs, affective needs, and cognitive needs, with due attention to environment, role and personality determining information seeking behaviour, and he adds:

> Many factors other than the existence of a need will play a part: the importance of satisfying the need, the penalty incurred by acting in the absence of full information, the availability of information sources and the cost of using them, and so forth (Wilson, 1981a:8).

He gives as examples of propositions to be tested:

1. Social role is a determinant of perceived needs for information.

* A biographical note on Drs H. P. Hogeweg-de Haart appears in *Social Science Information Studies*, 3 (3) on p. 147.

2. Roles which require the performer to generate new information or synthesize existing information are more likely to result in perceived information needs for their performers than other roles.
3. The level of social system at which a social role is performed will be a determinant of perceived needs for information (Wilson, 1981b:15).

Tamás Földi divides users of economic information (but this mig applied to users in other social science disciplines as well) according t fields of activity in

1. Economic research and education.
2. Economic policy-making.
3. Economic activity.

In a further division he distinguishes among

1. Decision-makers.
2. Those preparing decisions.
3. Practising or operative economists.

Combining these clusters this leads to nine groups with different inforn needs (Földi, 1976:518).

Nevertheless, Brittain has recently questioned the supreme information. He points to the tendency on the part of investigators of mation requirements to assume that information seeking and using paramount importance in all fields of research, administration and pr However, the general climate of opinion that information *per se* is good, th more we have the better, is changing. Brittain expects that:

> For a variety of political, social and economic reasons it is likely that the next decade will increasingly be concerned with the negative aspects of the information flow: the value of information will be questioned; the traditional approaches to the assessment of information requirements, particularly potential information needs will be challenged . . . (Brittain, 1981:89).

THE USE OF SOCIAL SCIENCE INFORMATION BY RESEARCI·

Information-seeking behaviour of social science researchers

Brittain reviewed the 18 user studies in social sciences that were availa the late sixties, but the only comprehensive one was INFROSS (Investic into Information Requirements of the Social Sciences), which was in it stage at the time. The earlier social science user studies drew heavily of much more numerous science user studies without taking into account tha

> The differences between the sciences and the social sciences make the methods of user studies in the one inappropriate in the other (Brittain, 1970:162).

The INFROSS Project was mainly concerned with the information re ments of social science researchers in various disciplines: anthrop economics, education, political science, psychology and sociology. methods were used:

1. A questionnaire circulated to a national sample from a population of all social science researchers.
2. 75 interviews with researchers and 50 interviews with practitioners.
3. Day-to-day observation of a very small number of social scientists (Line, 1971:413).

The results included the following:

1. Asked for the usefulness of methods of locating references 25 per cent of the respondents answered that they never used abstracts or indexes, library catalogues or book reviews, or searched on library shelves; 48 per cent never consulted librarians. Following up references in books and periodicals was found useful by 94 per cent and 59 per cent even rated this as the most useful method. Informal channels were quite commonly used for locating relevant references as well as for keeping abreast of new publications, and for keeping up with current research.
2. There are differences between disciplines in the use of abstracts. *Sociological Abstracts* was used by only 16 per cent of the sociologists, but *Psychological Abstracts* was used by 43 per cent of the psychologists, and *Geographical Abstracts* by 45 per cent of the geographers.
3. At the time the INFROSS questionnaire was circulated there were no published current awareness tools for the social sciences, nor a citation index. Nevertheless over half of the respondents replied that a citation index would be very useful.
4. Delegation of searching was rare: only 7 per cent delegated extensively. Users involved in applied or experimental research were more willing to delegate than those who wanted theoretical or conceptual material.
5. Asked whether any information problem had arisen during the current research, the most common special problem mentioned was physical access to, and availability of information (Line, 1971:417ff).

The main conclusions that emerged from the experimental information service that was set up at Bath University in the framework of INFROSS are that:

> Academic social scientists are not highly motivated to find references, nor well equipped in the means of doing so. Library catalogues are little used. Most formal information comes through reading primary journals and press' and publishers' notices of books. Much information is converged through informal channels of various kinds. Personal files are organized in highly personal ways (Bath University, 1972:101).

This is also confirmed by Stenstrom and McBride in a later survey of serial use in a social science faculty, in which the faculty relied heavily (50–70 per cent) on bibliographies and footnotes in journals and books to find references. They used the library only as a supplementary source of serial information, but they welcomed the idea of bibliographic services, such as listing titles of new journals, current contents, librarians' assistance in searching, and online searching (Stenstrom and McBride, 1979:428, 430).

Line takes it that:

> A good deal of the findings of INFROSS, as indeed of other user studies, seem to point strongly to the need for personal intermediaries.

With a large and complex body of knowledge on the one hand, and
on the other a user who needs the information but has insufficient
time and limited know-how, some means of linkage between the two
must be designed. If you bring the tools down to the user, you make
them ineffective, while attempts to take the user up to the tools have
not been outstandingly successful (Line, 1971:431).

*Information-seeking behaviour of social science researchers as compared with social
science and in the humanities*

In his overview of the DISISS studies Line states that in contrast to science

Many social scientists appear to be relatively little interested in
whether others are working in similar fields; conceptual work tends to
filter through gradually if it is of value, while experimental work is
most unlikely to be duplicated because the subjects under
investigation are almost certain to be different. In consequence, they
are less concerned than scientists if they are not up to date with
recent research or even if they are totally unaware of relevant past
research (Bath University, 1980:85).

Barbara Skelton compared the information-seeking behaviour of the
scientists of the INFROSS Project with that of scientists, drawing to
scientists on the studies of Herner (1954) and Flowers (1965), among or
Although the population of these studies is not quite comparable a
INFROSS population is composed mainly of academic researchers, whi
population of many science user studies includes technologists and scic
employed in a variety of environments, Skelton concludes that:

1. Within the formal system the main information sources are monog!
    and journals. Social scientists use both to an equal extent, but the s
    user studies show greater emphasis on journal literature.
2. Theses and patents are not important sources of information for e
    category.
3. Research reports tend to be slightly more important for scientists tha
    social scientists.
4. There is an indication that scientists regard informal personal contac
    more useful than do social scientists.
5. Government publications are used substantially by social scientists
    hardly at all by scientists.
6. Conferences are judged useful for informal personal contact, bot
    scientists and by social scientists, and not so much for the papers preser
7. Neither scientists nor social scientists consider reviews particularly u
    for locating information.

In summary Skelton notes:

The most heavily used methods by which scientists gain information
are personal recommendation, chance, abstracts/indexes and
citations. Social scientists regard citations, abstracts/indexes and
personal recommendation, in that order, as most useful (Skelton,
1971:16)

In Skelton's study the differences in information requirements betw
scientists and social scientists do not seem very strong. The reason is

though differences in information-seeking behaviour between disciplines are considerable, the most interesting points are lost in a generalized picture (Skelton, 1971:24).

This agrees with a comparison made by Ford of two studies among scientists, one among technologists and one among social scientists in which 38 per cent and 11 per cent of the scientists, 73 per cent of the technologists, and 17 per cent of the social scientists considered journals as their most important source (Ford, 1977:9).

Information requirements in the humanities and information-seeking habits of researchers in this field seem to have been scarcely explored, but Cynthia Corkill and Margaret Mann made a comparison between the needs of staff and PhD students in history, English, French, music and philosophy. They found that:

> The importance of libraries for research in the humanities was clear but it was equally clear that the use of libraries varied considerably according to the subject and the type of research being undertaken. While philosophers might need only their own university libraries and their private bookshelves, historians might use a whole range of libraries and other institutions (especially record offices) and their most important library might well be at a distance from their own university (Corkill and Mann, 1978:55).

Another conclusion of their study was that for both the social sciences and the humanities books and journal articles seem to be equally important, whereas for the sciences journal articles are much more important than books. But unlike the social sciences, and contrary to the sciences, information in the humanities does not readily go out of date (Corkill and Mann, 1978:55–56).

However, at a Seminar on Humanities Information Research, monographs were considered much more important than journals (Stone, 1980:92).

To summarize the similarities and the differences in information-seeking behaviour of the three categories, one of the conclusions of INFROSS might serve:

> The social scientist as an information user is at a point of a long continuum, from 'hard' to 'soft' social science. The 'hard' social scientist exhibits a number of characteristics that have been found in science user studies and is distinguishable in a number of ways from the 'softer' social scientist. The experimental psychologist and the econometrician are at the opposite poles from the sociological and educational theorist. One is tempted to suggest that the 'soft' social sciences are more akin to the humanities, but in fact some of the humanities are reasonably 'hard'; at least history and literature are concerned with historical data and documents, and with library texts, whatever interpretation may be placed on them (Line, 1971:430).

THE USE OF SOCIAL SCIENCE INFORMATION BY PRACTITIONERS

Introduction

According to Brittain the literature of pure and applied research in science finds little overlap, and communication channels are rarely shared. But in the

social sciences applied research is not such a clearly defined field (Br 1970:64).

There are a large number of practitioners who require, and sometimes use of, social science information. These practitioners may be social sci themselves (e.g., social workers, clinical psychologists, college of edu lecturers) or non-social scientists (e.g., architects, administrators, urba regional planners). They have little literature of their own and communi channels are not clearly defined. Hence many practitioners complain abc lack of useful social science information to solve their problems. Brittain as a possible reason that practitioners have neither the time nor the facili participation in the flow. There is a similar situation in science and techni In practice, however, applied scientists and technicians receive, wi assistance of professional communicators, information officers and keepers', the information relevant to their activities. Until recently the been few information officers in the social sciences (Brittain, 1970:123).

Maurice Line agrees with Brittain's view that practitioners have a sh of time to participate in the information flow, but he adds that they : general unawareness of information tools. In his opinion some inforr needs of practitioners could be solved by a more generous availabil existing reference works, and a greater awareness of the services local libraries can give. Keeping up with current practical trends presents problems, also because it is of little use to convey 'raw' research to practi in the form of journal articles and research papers. What they need is car prepared and temptingly presented packages, summarizing and evalu research findings that are sufficiently established for their practical implic to be clear. And since practitioners have an even stronger preferen informal communication than researchers, thought should also be giv developing channels of information.communication. Line concludes:

> Perhaps far too much effort has been devoted in the past to serving researchers, and far too little to serving the practitioners who are after all serving the community much more directly and obviously (Line, 1971:429).

The use of information by social workers

According to Maureen Webley:

> Social welfare personnel were able to function from day-to-day on the information they had; the few journals they read, and conferences and other informal methods of communication enabled them to keep reasonably up to date. However, it is now impossible to plan, carry out research or communicate internally or externally without drawing on a much more extensive fund of information (Webley, 1976:243).

Growing concern at the lack of provision of services for social workers . the establishment of a formal Working Group under the umbrella of the Social Sciences Group. The Working Group has pressed for the establish of some kind of centralized information bank in social welfare. In 1977 th to the establishment of *Social Service Abstracts*, issued by the Departme Health and Social Security Library. In the US *Social Work Research and A* (formerly *Abstracts for Social Workers*) had existed since 1964.

The DISISS Project included two evaluation studies in the field of social welfare, an analysis of the requests made to the National Children Bureau's question and answer service, and a user evaluation of an information service in social welfare, the *Wiltshire Social Service Staff Digest*. The analysis of the former showed that the majority (65 per cent) of enquiries were for information on a great variety of particular subjects, and less than 35 per cent of the users asked for references, research findings or statistics (Bath University, 1973b:8).

Of the interviewees among the readers of the *Wiltshire Social Service Staff Digest*, 99 per cent used factual information, 47 per cent used day-to-day information, 43 per cent claimed to use information on practical developments, but only 7 per cent said they made use of research findings in their work, and these were all directors, senior and professionally trained staff (Bath University, 1973a:24, 25).

The results of these evaluation studies make clear that the information requirements of a social service department can be met only to a very limited extent by the formal, non-specialized information systems.

In 1975 the three-phase Project INISS was begun. 'INISS' stands for 'Research project on Information Needs and Information services in local authority Social Service departments', and was an action-research investigation, taking account of the complexity of organizational structure and the different work-roles of staff members. Its point of departure was that:

> Whereas information staff usually draw on either the library custodial model (main function: to preserve recorded knowledge in retrievable form) or the scientific information service model (founded upon monitoring a well-developed system of scientific publication governed by strong pressure to publish the results of research) in the INISS project the focus is changed from the information officers' prescription based on their perception of the users' problems to self-prescription based on the users' perception of the users' problems (Streatfield *et al.*, 1981:244).

The conclusions of the project were:

1. Specialization of information services according to those work-roles where common needs can be discerned is desirable.
2. Information services should be integrated in the organizational communication patterns (Wilson and Streatfield, 1977:277; Wilson *et al.*, 1979:132).

In the Innovation Programme of Project INISS the effects of innovations are assessed in a six-step illuminative evaluation model, consisting of the following stages: value information, objectives, criteria for meeting the objectives, programme planning, programme operation, programme evaluation, and finally an evaluation of the evaluation process (Pritchard *et al.*, 1979:245-250).

Specific innovations were adopted for trial in one or more departments to see whether they worked in practice and what problems were to be overcome. These innovations were: training courses in information handling, book collections chosen by area office staff, a journal abstracting bulletin with a back-up service, an index of expertise, and the preparation of background papers, which all proved highly successful. A journal distribution, an information collection and a newssheet seemed to be less useful in the first run.

Conclusions of Project INISS Innovation Programme include:

38117

18

1. Field-based social services staff are deprived of published infor sources of all kinds.
2. Social services departments have sufficient 'readers' to just. development of services.
3. Instead of information specialists prescribing inappropriate service afar, it should be possible to create conditions in which these spe can share their expertise with social science practitioners at all or tional levels.
4. Taking information for granted should be stopped, relating inst availability more realistically to other aspects of organization, a tration and service delivery.
5. The only long-term strategy for social services departments is towards the creation of a department information policy and thu long-term benefits for social services clients (Project INISS, 1980;

In 1977 another three-year action research project, EISSWA (Exper: Information Services in two Social Welfare Agencies) started at W County Social Services Department. It owes in its conception much DISISS and INISS studies:

> Fundamental to the programme has been a policy of attempting to involve the practitioners themselves at all stages in the research process, from the initial identification of a particular information need, through the creation of systems and services intended to satisfy that need, to the evaluation and, if necessary, re-designing of those systems and services. The basic premise has been that practitioners will be more likely to use and—equally important—maintain information systems which they themselves asked for and helped to build than if those systems are imposed upon them from elsewhere (Blake *et al.*, 1979:276).

Next to more traditional SDI and question and answer services the res team has set up some experimental activities in which the workers them were involved, resulting in innovation and developments which are appar both simple and obvious and respond more to the needs of practitioners more sophisticated technical solutions.

It should be kept in mind, however, that in the hierarchy of needs of welfare practitioners reading and professional updating are:

> of lower priority than the more pressing problems of getting, organizing and sharing basic 'tools for the job' information: names of voluntary organizations, telephone numbers of GPs and health visitors, sources of financial assistance and the like . . . and it may be vital to find workable solutions to these 'lower level' problems before moving up the 'hierarchy of needs' to issues connected with reading and professional updating (Blake *et al.*, 1979:282).

The use of information by town and country planners

A comprehensive investigation into library resources and user requiremen town and country planning was made by Brenda White in 1969. Her samp 284 planners included 79 per cent practitioners and 21 per cent researchers teachers. Contrary to the researchers and teachers, who did not essentially in their information-seeking behaviour from researchers

teachers in other subjects, practitioners in planning had quite other preferences. They ranked maps first, followed by data from their own surveys, office records, statutory government publications, development plans, and advisory government publications. Journals, data from other surveys and statistical material ranked essentially lower (White, 1969:2.2.1).

As methods used by planners to locate information, Brenda White distinguished printed sources, library sources and personal sources. As to printed sources: the typical planner does not use abstracting or indexing services, or separately published bibliographies, but he finds references in books and periodicals. Library sources are considered to be quite useful, but he prefers personal sources. Brenda White states clearly what the importance is of the informal approach:

> The law of least resistance operates overall: the planner first consults his colleagues in his own department. From his colleagues, he may obtain the actual information which he wants; he may be referred to a journal, or a map, or a contact. But discussion with colleagues has an even more important function: it serves to crystallize the problem, to define the terms in which the question then can be put to a subject expert, or to a librarian . . . The informal approach thus serves as a preliminary to a formal approach (White, 1969:2.2.2.4).

She recommends three ways in which the provision of planning information can be improved:

1. Improving the physical availability of sources in planning organizations and in general libraries.
2. Making planners aware of the wide range of sources which exist.
3. Providing planners with packaged concentrated information which is supplied directly to them and which can be immediately applied to practical situations.

The INFROSS Project showed that users are usually unable to state what improvements might be made in information services in order to satisfy their needs more fully, in the absence of an actual service to which they can refer. It is therefore necessary to create new services which can be tested in practice. Because of this the DISISS team in collaboration with *Geo Abstracts* has set up a series of experimental issues for planners, in which various parameters could be tested. From the results of the experiment Arms concluded:

> It appears that frequency is not a key factor, but that the size of the issues was definitely important. The index in a current awareness service did not prove its value. There appeared to be a balance between preferences for abstracts and for simple bibliographic citations. This seems to suggest that there are two types of user, one wanting a very simple service with small coverage, no index and providing titles only, another preferring a fuller service (Arms, 1976:12).

On the whole academics want coverage as wide as possible, while practitioners are unwilling to waste time following up items which are difficult to obtain and would prefer more compact issues (Bath University, 1974:72). This raises the question whether one service could serve both researcher and practitioner.

20

The use of social science information by policy-makers

In the early sixties research into the structure and the role of social information was started with research into the flow of social information. In the late sixties and early seventies information requirem social science researchers and practitioners and their use of information into the limelight; in the middle seventies also the use of social information by non-social scientists became an important subject of re and especially the role of social science research in policy-making and de making.

Paul F. Lazarsfeld and Jeffrey G. Reitz in assessing the interaction b policy-making and sociological research state:

> One view of decision-making assumes that once goals, interests, and normative considerations are spelt out and all the facts analyzed, decisions inexorably emerge. This is as mistaken as the opposite view that rational analysis invariably impedes the 'art' of decision making. A more balanced approach suggests that, although systematic knowledge can contribute to decisions there is always a gap between knowledge and recommendation (Lazarsfeld and Reitz 1975:98).

They distinguish two 'gap situations'. In one, research has pr material which seems relevant to a policy problem, but where the av knowledge is still too general and rather amplifies the difficulties t decision-maker by showing the complexity of the situation. The inforr still needs to be boiled down to the choice between a few possible cour action. This makes for the second type of gap situation, in which the dec maker has to make a qualitative cost-benefit analysis (Lazarsfeld and 1975:101).

Though information, if boiled down to clear alternatives, is important : decision-maker, it seems to be only one of the elements that under decisions. His interests and values are other elements in the process. Th of information seems to be more that of giving certitude in negotiation that of answering to cognitive needs.

Tamás Földi stresses the point that the information needs of decision-n differ greatly from those of research workers, teachers, and also of engaged in preparing decisions at lower levels:

> The decision-making process involves a supply with sufficient data in due time and for the right subject to decision-makers. The latter have no opportunity to control, select and analyse the data emanating from scientific information. Considering all these factors, information services have to comply with the following information needs for decision-making: complexity, selectivity, conciseness, reliability, supply in due time (Földi, 1981:2).

James Coleman notices that in political science little attention has been to information as a basis for political decisions:

> Political theory and constitutions pay little attention to questions involving information: how legislative bodies obtain information to guide policy; how voters obtain information about office-holders and their challengers; how information comes to shape the very law in which policy questions are framed; how the resolution of conflicting

interests is affected by the available information (Coleman, 1976:304).

In *Information for action; from knowledge to wisdom* 15 authors discussed H. G. Wells' 'world brain' idea in a set of principles to a *World Information Synthesis Encyclopedia* (WISE). In the conclusion of this book it is stated that:

> Large quantities of data, even if highly organized for processing by computer and rapidly accessible may bewilder more than help . . . Research on and development of information systems have tended towards improved access to scientific and technical documents by specialists in order to keep the communication system within the scientific community in good health. Though that is important it does not compare with the seriousness of the day to day problems faced by community leaders and individuals as non-specialists, and with which information systems could perhaps be expected to help. Much of the knowledge and understanding they need is not scientific (Kochen, 1975:152).

The WISE system should direct the user to potential technical experts and also to potential participants in the decision-making process.

Härö points to the lack of communication between the producers and the users of information:

> The key problem in developing information-services which are really to serve their purpose is to answer the question: what should be known? Too little emphasis has been given to laying responsibility for information content primarily on decision-makers, not on statisticians or data processors. Thus statistics and generally quantitative information tend to be undervalued and there is no very close cooperation between bureaucrats and researchers, especially theoretically oriented ones (Härö, 1977:430).

A quite comprehensive investigation on the use of social science information by policy-makers was executed by Nathan Caplan and his team at the University of Michigan. Their study on the nature and degree of social science information utilization in decision-making among 204 upper level government officials showed that there is:

> Evidence of a high level of interest in and receptivity to social science information inputs in policy-level decisions, but little evidence that such information, either directly or indirectly, enters into decision-making activity (Caplan, 1975:50).

Caplan and his team were able to identify 575 examples of social science use among the 204 interviewees; 385 involved the use of primary research sources, rather than research information from secondary sources such as reviews or digests; 60 per cent involved the use of conceptual information, 40 per cent of instrumental information, i.e., the application of more narrowly defined or purpose-specific, usually data-based information. Sources of social science information of the respondents were: newspapers (81 per cent), government reports (81 per cent), staff assistance (76 per cent), books (76 per cent), professional journals (70 per cent), magazines (69 per cent), TV and radio (50 per cent (Caplan, 1975:8, 9).

The value of social science inputs to policy-makers in terms of its usefulness in various policy-related areas, arranged in order of perceived usefulness was:

1. Sensitizing policy-makers to social needs.
2. Evaluation of ongoing programs.
3. Structuring alternative policies.
4. Implementation of programs.
5. Justification of policy decisions.
6. Providing a basis for choosing among policy alternatives (Caplan et al., 1975:22).

Given the importance of sensitizing policy-makers to social proble given the repeatedly expressed desire of respondents that there should lag in their awareness of social reality, Caplan concludes:

> One of the important functions of social science research should be to inform and educate policy-makers to see contemporary social reality as it is and to monitor social change over time (Caplan, 1975:64).

A multivariate analysis of respondents' reactions to some 36 representing a set of the most common non-utilization theories (orde three categories: knowledge-specific theories, two-communities theors policy-maker-specific theories), resulted in a predominance of th communities theories to account for the largest proportion of ex variance (Caplan, 1975:57).

In these two-communities theories it is argued that the social scientu policy-makers live in separate worlds, with different and often con values, different reward systems, and different languages. Responden three reasons for deliberate non-utilization of social science information:

1. The data are not believable on grounds of objectivity.
2. The findings are counter-intuitive.
3. The data are valid but rejected because they are politically unfeas use (Caplan, 1975:59).

Over 70 per cent of the respondents agreed, or strongly agreed, wi statement:

> The non-use of social scientific knowledge is an example of the general principle that people tend to trust their intuition about social problems more than anything else (Caplan, 1975:60).

Although Caplan and his team did not include a set of items comp utilization across the sciences in their interviews, Caplan believes that:

> The social sciences probably suffer no more from non-use than other sciences in the formulation of policy at the national level, . . . and, further, that if some sort of cost-benefit comparison were made on the utilization per dollar spent, the social sciences would compare favourably with other sciences (Caplan, 1975:62).

Oliver C. Moles describes an endeavour to couple scientific know production to policy goals and objectives in the Social Science inform Service (SSIS), sponsored by the Society for the Psychological Study of S Issues (SPSSI), which is a division of the American Psychological Associat

> The Service seeks to inform legislators and their staffs as to the likely social and psychological implications of various legislative proposals where there is applicable evidence from social science research. A basic goal of the Social Science Information Service is to reduce the

gap in the amount and quality of social science information currently available to the Executive and Legislative branches of government (Moles, 1976:228).

SSIS is particularly interested in bills still being formulated, but also in bills already submitted and in the legislative process. Nationally known social scientists are asked to analyse the problem and determine what social and psychological phenomena are likely to be subsumed within it. A computerized literature search will be done to obtain information about probable persons who might be able to provide an analysis of what phenomena are actually involved, to provide the latest findings or suggestions of who are the most knowledgeable persons in the area, and to give references to relevant unpublished work (Moles, 1976:229).

Among the subjects on which SSIS presented comments and research findings, according to Moles, are the problems of ageing, nutrition, national health insurance, work satisfaction, juvenile delinquency and non-economic effects of the recession.

SSIS strongly opposes the charge of being a lobby of some sort. This being so, it is an interesting and unique approach to information, because it takes the initiative instead of waiting for requests, as most information services do.

This mode of approach might constitute an exception to Coleman's view that:

> Whereas social policy decisions are timed in accordance with legislative sessions, social research tends to be a lengthy process and often does not provide results in time to be of any use for policy making. Therefore social research is less frequently used by an administrative body with authority to legitimate a policy than by external groups, without authority, to challenge existing policy (Coleman, 1978, according to abstract in *Library and Information Science Abstracts*).

It might really need a whole scale of various methods, orthodox and unorthodox, to convey up-to-date and ready-to-use social science information for the benefit of policy-makers, other than what might be called information for a 'social science perspective' on the questions involved.

However, the problem is in the limelight, judging also by the OECD publication *Social sciences in policy making*, prepared by an Expert Group on Social Sciences on behalf of the Committee for Scientific and Technological Policy. It recommends strengthening communication between governments and the social sciences in order to identify which problems are amenable to research and likely to yield results relevant to policy-making, and strengthening existing mechanisms where they have been found insufficient to influence decision-making (*Social sciences in policy making*, 1979:51).

It is recognized in the report that the development of new and more effective approaches to utilization depends on the extent to which policy-makers are prepared to experiment in the application of social science findings. It recommends therefore that:

1. Parliamentary bodies should expand and improve their capabilities for taking social science findings into account.
2. Closer links with social science research between local, regional and national levels of the administration and in all spheres of the political process need to be developed (*Social sciences in policy making*, 1979:52).

CONCLUSIONS

1. In social science there is a marked difference in information-se behaviour between disciplines, but the same applies to science and humanities. Taking the sciences and the social sciences in general information-seeking behaviour of scientists and social scientists doe seem to differ more than that among the social science disciplines t selves.

2. Investigations in social science information started in the early s with research into the flow of social science information. From the sixties onwards the focus of interest shifted towards the informa requirements and information-seeking behaviour of social sci researchers, followed by investigations and experiments on the infor tion requirements and information-seeking behaviour of social scien in non-research environments and of non-social scientists in need social science information. These investigations have shown that the a considerable difference in information requirements and use information between the three groups. Also within these groups, u may have different requirements according to their activities.

3. Until quite recently social science information services have been conc trating on the information requirements of researchers. Researche however, do not use these services to a great extent, but prefer referen in the literature and informal contacts for their literature searches. The rarely delegate their searching. Nevertheless there seems to be a need f personal intermediaries to show the way in the information jungle.

4. Practitioners seem to lack time to participate in the information flow an seem to have little awareness of the existing information tools; moreov they have other information requirements than the available informatio tools can offer. Information specialists should therefore, in collaboratio with the practitioners, design special tools, answering to the practitioners requirements and in accordance with the hierarchy of their needs.

5. Though non-social scientists requiring social science information, such a policy-makers, show a high level of interest in this information, it seems to serve more to sensitize them to social needs than to provide the basis for their decisions. It seems to need unorthodox methods to serve policy-makers with up-to-date and ready-to-use information for their decisions.

6. It is to be expected that in the near future the negative aspects of information will feature more strongly in library research.

The implications of these conclusions would appear to be for information services:

1. To enhance user-friendliness of information systems.
2. To cooperate in order to be able to offer a greater variety of types of information, such as short bibliographical references, indicative and informative abstracts, reviews, overviews of specific data, and evaluations of research findings.
3. To cooperate in order to be able to specialize in serving specific user groups.
4. To ameliorate retrieval systems to reduce unspecific and redundant information.

The implications for information workers would appear to be that they should try to:

1. Look at information problems more from the user's point of view than from their own.
2. Be aware of the relative value of information in the hierarchies of needs of researchers as well as of practitioners and policy-makers.
3. Help specific groups of users with common needs to develop tailor-made services to answer their specific information requirements.

REFERENCES

ARMS. W. Y (1976). Individual secondary services and their optimization. In: *International Workshop of Secondary Service Producers*, pp. 9–13. (BLR & DD Report No. 5289)

BATH UNIVERSITY (1972). *Experimental information service in the social sciences 1969–1971; final report*. Bath University, Bath University Library.

BATH UNIVERSITY (1973a). *User evaluation of an information service in social welfare: report of an interview and questionnaire study of the Wiltshire Social Service Staff Digest*. Bath University, Bath University Library. (DISISS Research Report, Series B, No. 1)

BATH UNIVERSITY (1973b). *Analysis of requests to the National Children's Bureau's question and answer service*. Bath University, Bath University Library. (DISISS Research Report, Series B, No. 2)

BATH UNIVERSITY (1974). *The relationship between usefulness and style of secondary publication; an experimental information service for planners*. Bath University, Bath University Library. (DISISS Research Report, Series B, No. 3)

BATH UNIVERSITY (1980). *Towards the improvement of social science information systems; overview of research carried out 1971–1975*. Bath University, Bath University Library. (DISISS Research Report, Series A, No. 1)

BLAKE. B., MORKHAM. T. and SKINNER. A. (1979). Inside information: social welfare practitioners and their information needs. *Aslib Proceedings*, 31, 275–283.

BRITTAIN. J. M. (1970). *Information and its users; a review with special reference to the social sciences*. Bath: Bath University Press.

BRITTAIN. J. M. (1981). Problems of establishing information requirements in the social sciences. In: *Dux femina facti; aspecten van social-wetenschappelijke informatie* (E. Z. R. Cohen *et al.*, ed.), pp. 77–89. Amsterdam/Oxford/New York: Noord-Hollandsche Uitgevers Maatschappij.

CAPLAN, N. (1975). The use of social science information by federal executives. In: *Social research and public policies* (G. M. Lyons, ed.), pp. 46–67. The Dartmouth OECD Conference. The Public Affairs Centre. Dartmouth College, Hanover, New Hampshire.

CAPLAN, N., MORRISON. A and STAMBAUGH. R. J. (1975). *The use of social science knowledge in policy decisions at the national level; a report to respondents*. Ann Arbor: University of Michigan.

COLEMAN. J. S. (1976). Policy decisions, social science information, and education. *Sociology of Education*, 49, 303–312.

COLEMAN. J. S. (1978). *The use of social science in the development of public policy*. Paper presented at the Annual Meeting of the American Educational Research Association (Toronto, Ont., March 27–31, 1978).

CORKILL. C. and MANN. M (1978). *Information needs in the humanities: two postal surveys*. Sheffield: University of Sheffield, Centre for Research on User Studies. CRUS occasional paper 2. (BLR & DD Report No. 5455)

FLOWERS. B. H. (1965). Survey of information needs of physicists and chemists. *Journal of Documentation*, 21, 83–112.

FÖLDI. T (1976). Economic information: demand and organization. *International Social Science Journal*, 28, 518–530.

FOLDI, F. (1981). *Scientific information in Hungary in the service of economic summary.* FID/SD Discussion Paper 2/1981, p. 2.

FORD, G. (1977). *User studies: an introductory guide and select bibliography.* Si sity of Sheffield, Centre for Research on User Studies. CRUS oc (BLR&DD Report No. 5375)

HARO, A. S. (1977). Strengthening health information services. *Internati Journal, 29*, 419–432.

HERNER, S. (1954). The information gathering habits of workers in pure science. *Industrial and Engineering Chemistry, 46*, 228–236.

KOCHEN, M. (ed.) (1975). *Information for action: from knowledge to wisdom.* Francisco/London: Academic Press.

LAZARSFELD, P. F. and REITZ, J. G. (1975). *An introduction to applied socio.* London/Amsterdam: Elsevier.

LINE, M. B. (1971). The information uses and needs of social scientists INFROSS. *Aslib Proceedings, 23*, 412–434.

MERSEL, J., DONOHUE, J. C. and MORRIS, W. A. (1966). *Information tran research.* Sherman Oaks, Cal.: Informatics Inc.

MOLES, O. C. (1976). The Social Science Information Service. *Policy Sc. 228–234.

PRITCHARD, J. H., STREATFIELD, D. R. and WILSON, T. D. (1979). An illumina programme for assessing information services. *Aslib Proceedings, 31.*

PROJECT INISS final report on the innovation programme, 1978–1980. (1 Sheffield University, Department of Information Studies.

SKELTON, B. (1971). Comparison of results of science user studies with into information requirements of the social sciences'. Bath: Bath (DISISS Working Paper, No. 1)

Social Sciences in Policy Making (1979). Paris: Organisation for Economic (and Development.

STENSTROM, P. and McBRIDE, R. B. (1979). Serial use by social science facul *College and Research Libraries, 40*, 426–431.

STONE, S. (ed.) (1980). *Humanities information research: proceedings of a semina* University of Sheffield, Centre for Research on User Studies. CRU paper 4. (BLR&DD Report No. 5588)

STREATFIELD, D. R., WILSON, T. D. and CLARK, D. (1981). A problem-ba approach to information for practitioners. *Libri, 31*, 243–257.

WEBLEY, M. (1976). The Aslib Social Sciences Working Group on so information: a review of its development and current activities. *Asli 28*, 243–247.

WHITE, B. (1969). Information provision in town and country planning report. London: The Library Association. (OSTI Report No. 5059)

WILSON, T. D. (1981a). On user studies and information needs. *Journal of D 37*, 3–15.

WILSON, T. D. (1981b). Sociological aspects of information science. *Internation Information and Documentation, 6*, 13–18.

WILSON, T. and STREATFIELD, D. R. (1977). Information needs in local autho services departments: an interim report on Project INISS. *Journal of Do 33*, 277–293.

WILSON, T. D., STREATFIELD, D. R. and MULLINGS, C. (1979). Information nee authority social services departments: a second report on Project INISS *Documentation, 35*, 120–136.

Search strategies: some general considerations

The subsequent chapters in the *Manual of Online Search Strategies* discuss in detail the strategies that can successfully be used to retrieve digitized information in a wide range of subject areas. However, more general topics common to all areas are assembled in this opening chapter, rather than being duplicated throughout the book. The objective here, then, is to provide introductory comments on the users of information systems – the information seekers – and the variety of technologies that can now be used to offer digitized information to those seekers. It also reminds seekers that the same database may differ in content or structure from platform to platform or even from vendor to vendor. Subsequent chapters discuss the specific indexing characteristics of many databases; in this chapter some general points are made about the relative merits of searching on assigned controlled terms as against natural-language terms as found in the database documents themselves. The interface to any retrieval system is a crucial determinant both of user satisfaction and success, so interface design criteria are very briefly reviewed. Finally, the chapter discusses the important topics of search and database evaluation.

Information seekers

There is a tendency to discuss information seekers as if they are a homogeneous group. In reality, of course, each seeker is an individual who brings to the workstation a particular set of personal characteristics, subject knowledge and retrieval skills as well as a unique information need; all of these influence the search outcome. Nevertheless, it is both practical and useful to sort them into categories according to certain broadly defined characteristics.

The most common characteristic relates to the seekers' level of experience in information retrieval. Applying this measure, seekers can broadly be categorized as novice or experienced. Unfortunately, no generally accepted criteria have been formulated to assist in this distinction. Borgman (1996) suggests that an information seeker requires three layers of knowledge:

- conceptual – to convert an information need into a searchable query
- semantic – to construct a query for a given system
- technical – to enter queries as specific search statements.

An experienced searcher, therefore, would be someone possessing such knowledge and able to implement these three actions. This begs the question, of course, as to *how* such tasks might be assessed and judged as well or badly performed. Hsieh-Yee (1993) offers more specific criteria: novice searchers are non-professional searchers who have little or no search experience and have not taken courses in online searching or attended relevant workshops provided by librarians or system vendors; experienced searchers are professional searchers who have at least one year of search experience and have either taken courses on online searching or attended workshops provided by system vendors. This definition, equating novice with non-professional (or end user) and experienced with professional, suggests that only the latter can become experienced. While user studies do indicate that many non-professionals are not especially effective when searching, it is a sweeping statement to suggest that only professional searchers can attain expertise, and that this is achieved only by taking courses or attending workshops. Extensive information seeking on the World Wide Web, and discussion of this activity in popular magazines and on radio and television, is providing a level of familiarity (if not always expertise) with information retrieval among diverse user groups regardless of formal instruction. Despite the definitional problems, the terms 'novice' and 'experienced' recur in discussions of information seeking.

A related distinguishing characteristic is between an information professional (or information intermediary) – the person who conducts a search on behalf of a client – and an end user – the person who actually wants the information to answer a specific need (as we have just seen, Hsieh-Yee uses this characteristic to distinguish between the novice and the experienced). This distinction was especially valid when most online searching was conducted by intermediaries – librarians or other information specialists – rather than by the actual information requester. Information professionals were considered experienced users, as was generally the case, and end users were novices who conducted searches rarely and with little or no preliminary training (often also true). Much searching is now undertaken by end users – a consequence in large part of simpler retrieval interfaces and a wider range of accessible information on CD-ROM and the World Wide Web.

Information seekers can also be categorized by a variety of other criteria. Do they have a thorough knowledge of the subject in which that search is to be conducted (whether novice or experienced, end user or intermediary)? The subject specialist's search is likely to be different from the non-specialist's because, for example, the former will have a greater awareness of the subject's terminology, and therefore be better placed to select suitable search terms. The specialist should be better at selecting the most suitable sources for the search, which has become increasingly difficult as electronic information resources proliferate. The

specialist should also be able to judge the relevance of retrieved information and adjust a search strategy if this seems appropriate.

Many user studies have investigated young adult information seekers in the context of a university library. Most researchers are located in university departments, and the most obvious and accessible subjects for their studies are to be found on their own doorsteps. It would be difficult to argue, however, that university undergraduates necessarily represent a cross-section of information system users in general. More recently, greater interest has been shown in other user groups. A prime example is children, who increasingly use online information systems both in school and from their homes. Do systems that have been designed for adults work just as effectively for children, or do the different cognitive skills and knowledge bases of children demand information systems that have been specially designed with this specific user group in mind? The same might be said of users at the opposite age spectrum. Elderly citizens are likely (for the time being, at least) to be less familiar with computers, to have poorer eyesight and less precise hand movements than their juniors. Should this make a difference, for example, to the kinds of interfaces provided by the OPAC or Web site?

Seekers might also be differentiated by their search objective: are they trying to find absolutely everything that is available on a topic, even if only tangentially linked to it, or only a little information directly on the topic? These distinctions will almost certainly affect the search strategy, and perhaps the choice of database.

Finally, psychological factors such as attitude, motivation and cognitive style can differentiate users (or even the same user on different occasions), although attempts to measure their impact on search outcome have proved far from conclusive. As in other types of human performance, it is extremely difficult reliably to isolate individual characteristics that can then be tested for an effect on searcher performance. Fidel and Soergel (1983), for example, identified over 200 variables that could come into play when investigating searching.

Some studies have failed to find a clear, positive relationship between search experience and search results. Lancaster *et al.* (1994), for example, compared CD-ROM searches on a bibliographic database by graduate student end users and skilled university librarian intermediaries. The librarians were able to find, in total, more relevant records on the database than the students, but a higher percentage of the records retrieved by the students were judged relevant by them. The greatest problem encountered by the students was failure to identify and use all the terms needed to perform a more complete search; they were less successful in identifying synonyms than the librarians. Hsieh-Yee (1993) found, however, that search experience positively affected search behaviour, especially when the experienced searchers had some subject knowledge relevant to the topic of the search. They used more synonyms and tried more combinations of search terms than novices.

Many studies have commented on the positive evaluations typically made by end users of their own search results, and questioned whether such optimism is really justified (see, for example, Lancaster *et al.*, 1994; Martin and Nicholas, 1993). Sanderson (1990) considered that no matter how user-friendly the system,

end users need clear directions to help them get the best results; training programmes should emphasize system capabilities and the kind of information that can be obtained, and should include hands-on sessions in which users are taught how to do basic searches.

Technologies

Digital (or electronic) information can now be found using several related, but distinct, technologies. Remote online information systems, accessible via dial-up telecommunication networks (typically only to users who have signed a contract with the system) remain important purveyors of databases. Examples of such systems are DIALOG, DIMDI and STN. The first public demonstration of such an interactive retrieval system was made by the System Development Corporation in 1960, and until the late 1980s these online systems dominated the digital information market.

In the 1980s libraries began to replace their card catalogues with Online Public Access Catalogues (OPACs). Unlike the traditional online systems, whose use was largely confined to information professionals, OPACs were intended for all library users. Another development of the late 1980s was the CD-ROM, an optical rather than a magnetic data storage medium. Although data could not be deleted from or added to a CD-ROM (as is the case with magnetic storage media), the CD-ROM proved to be a cheap and efficient medium for publishing digital information, thereby extending the market from the institutional to the domestic setting. Increasing numbers of CD-ROMs are purchased to be used on home-based personal computers for recreational purposes (in many cases the CD-ROMs contain games rather than 'information' *per se*).

Technology developments have continued into the 1990s. In their early days, CD-ROMs could not be networked, or if this was possible then response times were severely degraded. Institutional exploitation of CD-ROM technology was greatly facilitated by the emergence of networked versions of many CD-ROM titles. CD-ROMs also have been joined by related optical storage devices that greatly extend the quantity of data that can be stored on a single disc (the Digital Video Disc/Digital Versatile Disc or simply DVD-ROM, for example, can accommodate around seven times more data than a CD-ROM; data can now be added to a disc so it is not just read-only (the introduction of CD-R – CD Recordable – for example, allowed institutions or individuals to create their own CDs and therefore to store locally created data on this medium).

Undoubtedly, the most dramatic development of the 1990s, however, has been the rapid growth of the Internet, and especially the World Wide Web that makes statistics outdated by the time they are collected. In late 1999, however, one Web search engine – ALTAVISTA – claimed to index 250 million pages (Notess, 2000) while INKTOMI, in January 2000, had over 1 billion documents in its database, each relating to a unique page (Inktomi, 2000). And not even the largest engines are able to index anything like the entire Web.

Web search engines can either be general, such as ALTAVISTA <URL http://www. altavista.digital.com/> or NORTHERNLIGHT <URL http://www. northernlight.com/>, attempting to provide access to the Web as a whole, or specialized, such as AL IDRISI <URL http://www.alidrisi.com> (in this case, Arabic-language pages) or WAITER.COM <URL http://www.waiter.com/cgi-bin/SCMMOS/RegSys/AutoRegHome.cgi> dealing with take-out food delivery. An increasing number of search engines cover a specific country or region, such as SEARCHUK <URL http://www.searchuk.com/> or NZ EXPLORER <URL http://nzexplorer.co.nz/> (for over 250 000 Web pages in New Zealand).

At the other extreme, meta search engines like DOGPILE <URL http://www. dogpile.com/index.html> or METACRAWLER <URL http://www.metacrawler.com/ index.html>, search simultaneously on multiple regular search engines (useful because no single engine, including the very largest general ones, in practice indexes more than a part of the entire Web). Hock (1999) advises that these meta-engines are most useful when searching for a single, very rare word or when it is not important that all the relevant records are found (because most of the meta-engines only return between ten and 30 pages from each target engine, and do not employ sophisticated search syntax such as Boolean term matching, even if the user enters them).

Web search engines should be distinguished from Web directories, the best known of which is YAHOO! (see Figure 1.1). These directories provide hierarchical menus of subjects that can be used to narrow a search, but will only give access to a fraction of the Web.

The Web has made digital information an everyday fact of life for millions of people across the globe. In part, it has provided an alternative platform for the kinds of databases that previously were only found on traditional online systems, OPACs or CD-ROMs. But it has also extended the type of information that can be accessed digitally by enabling practically any institution or individual to create a Web site from which information can be disseminated around the world. Virtual libraries such as the AUSTRALIAN BIOLOGICAL RESEARCH NETWORK (ABREN) Virtual Library <URL http://abren.csu.edu.au/abren/library/Organisation.html> and gateways such as the SOCIAL SCIENCE INFORMATION GATEWAY (SOSIG) <URL http://sosig.ac.uk/> are just two examples (see Figures 1.2 and 1.3).

It is interesting to see these changes mirrored in the various editions of the *Manual of Online Search Strategies*. The first edition, appearing in 1988, was confined to traditional dial-up online systems. By the second edition, in 1992, CD-ROMs also occupied a prominent place in most of the chapters. The most casual perusal of this third edition will reveal the central role now being played by the World Wide Web, alongside traditional online systems and CD-ROMs.

Many databases are now available on more than one technology – dial-up online system, CD-ROM and the Web, as well as other possibilities such as diskette or magnetic tape – and from more than one supplier. For example, the MEDLINE database is found on several online systems (including LEXIS-NEXIS, OCLC FirstSearch, Ovid Online, DIALOG, DataStar, and STN), as CD-ROMs from, for example, SilverPlatter, and in at least two Web versions (PUBMED
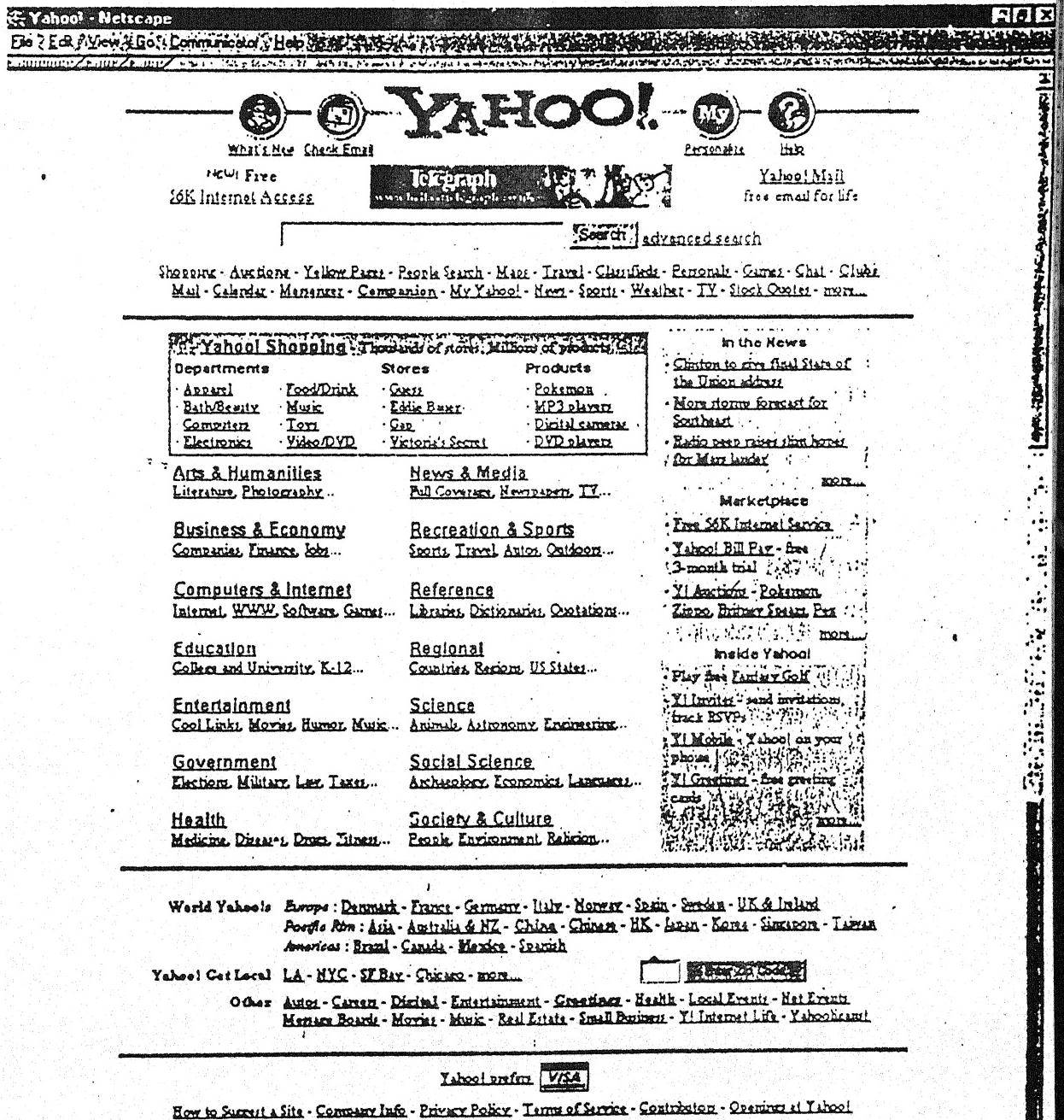
Figure 1.1   Opening directory on YAHOO!

and INTERNET GRATEFUL MED), as well as being available on tape for local installation. Some database producers, such as the National Library of Medicine, have long made their products available themselves over dial-up routes or on CD-ROM; the Web has encouraged many more, like the Institute for Scientific Information (ISI), to follow suit.

When a database is available via several technologies, how should a choice between them be made? Both the traditional online systems and the World Wide Web require the use of a data transmission network to connect the user's workstation with the database server. Response times can be variable depending
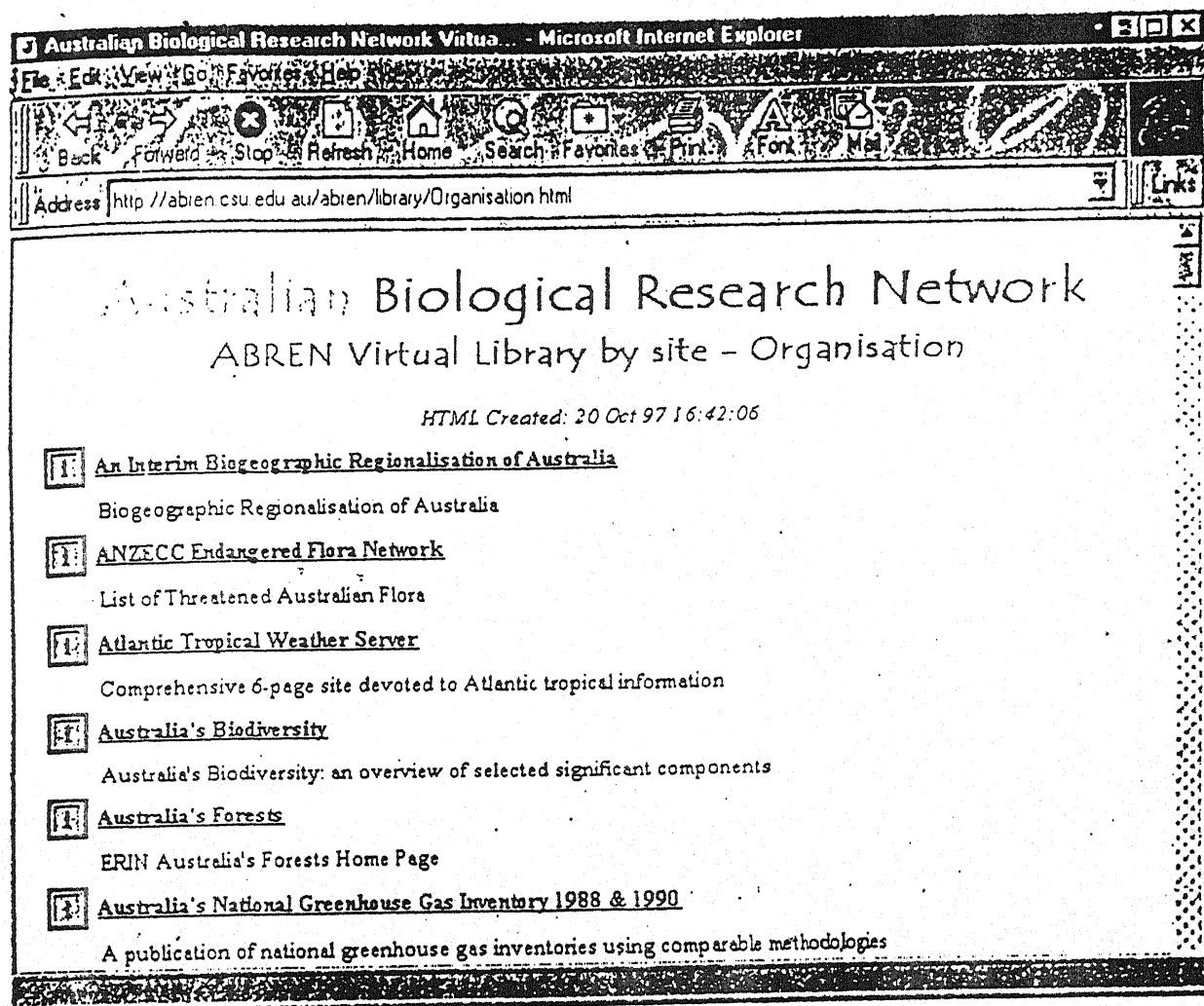
Figure 1.2 Excerpt from the opening screen of the AUSTRALIAN BIOLOGICAL
RESEARCH NETWORK

on network use (for example, from Europe the use of North American-based hosts
tends to be faster in the morning – when most North Americans are in bed rather
than hunched over their computers – than in the afternoon or early evening). The
use of graphics on Web versions of databases can enhance database content
compared with traditional online versions, but also slow down data transmission.
CD-ROMs and leased tapes (as well as OPAC searching within the OPAC's home
library) normally eliminate the need for long-distance data transmission.

Although interface dialogue modes – command languages, menus of various
kinds, and direct manipulation or an object-oriented interface (an interface that
provides a visual environment for the dialogue between user and computer) – are
not strictly related to individual technologies, in practice command searching has
been associated with traditional systems, whereas menus and direct manipulation
techniques have been more common with CD-ROMs and now the Web. Vendors
such as The Dialog Corporation do offer menus on their traditional online version
as well as commands, and commands on their Web version as well as menus, but
experienced command-mode users may still find the online version preferable to
the Web version because it is easier to view the search developing in a linear

Social Science Information Gateway

SOSIG is an online catalogue of thousands of high quality Internet resources relevant to social science education and research. Every resource has been selected and described by a librarian or subject specialist. Use the buttons at the top of each SOSIG page to search or browse the catalogue.

- About SOSIG
- SOSIG Contacts
- SOSIG Training
- European Collaboration
- Social Science Departments
- Other SOSIG Services

Figure 1.3   SOCIAL SCIENCE INFORMATION GATEWAY (SOSIG)

sequence (although it must be conceded that familiarity also plays a big part – searchers who have learned their skills in one mode will tend to cling to it through habit and familiarity).

Pricing structures differ between the various platforms, which in turn can encourage somewhat different search strategies. Traditionally the online systems charged largely according to connect time – the duration of a search from an initial system log-on to a final system log-off. This pricing structure placed a premium on short searches, which in turn favoured experienced, professional searchers who could quickly find the desired information, and a command-driven interface which in the hands of such an experienced searcher is faster than a menu-driven or even a direct manipulation interface. In contrast, CD-ROMs incurred no incremental costs when a search was undertaken. Like a book or a serial, the CD-ROM was obtained either by direct purchase or by subscription independently of the number of times used. CD-ROMs were therefore more hospitable to novice or occasional searchers and could employ user-friendly menus even if these proved much slower than commands. Although Web searching does involve a connect charge (in the form of a charge for an agreed number of hours connect time over a given period of time – typically a month), this is so small in practice as to be largely or entirely discounted, and therefore, as with CD-ROM searching, professional expertise is

less needed. In the case of certain vendors that offer Web services, such as DIALOG, a connect charge is incurred to the vendor as well as the Internet provider. Such connect time, however, is only measured when the host computer is actually occupied with the search; as the client workstation logs on and off the host server, the counter is switched on and off, thus allowing the searcher to pause for thought within a search at no connect cost – a luxury never afforded by the traditional online systems.

A major drawback with the CD-ROM is that its data cannot be updated except by issuing a new disc. A new CD-ROM disc cannot really be updated more frequently than every three months or so. This poses no problems for databases containing rather static information that does not require frequent updating, such as encyclopedias. Where more regular updating is required – monthly, weekly, daily or even real-time – online and Web services have a clear advantage. For example, someone searching to establish whether a particular invention has already been patented will want to ensure that the database is as up-to-date as possible (Volume II, Chapter 2). For this reason, a CD-ROM may be searched for retrospective information, but its online equivalent searched for the most recent updates. Increasingly, hybrid CD-ROMs are appearing that allow an initial disc search to be updated from a dial-up online or Web connection. Examples include Microsoft's ENCARTA and World Books' MULTIMEDIA ENCYCLOPEDIA.

Dial-up online systems and CD-ROM search engines typically offer similar options: Boolean term matching, stem truncation (usually only right-handed – masking the end of the word so that all words beginning with that word stem will be found – although DIMDI also offers left-hand – to find all words *ending* in the search term), embedded truncation (to mask letters within words in order, for example, to find both 'woman' and 'women'), adjacency (or proximity) searching (to locate words next to or close to each other as phrases), and field searching. Increasingly, the Web search engines are also providing these features, although fields are much less formally defined on a typical Web page. One advantage that remains with the online systems is the ability to store and re-use if necessary sets retrieved earlier in the search, although some Web interfaces to an existing online service (for example, Ovid) also now offer this option. The Web search engines offer an additional feature – ranking of retrieved documents. Whereas the online systems and CD-ROMs have normally displayed any retrieved data records in chronological sequence – the last ones to have been added to the database in any retrieved set will be the first displayed on the screen – Web engines attempt to rank by probable relevancy (to the initial query) the most relevant retrieved documents being displayed first. Such ranking typically relies on various techniques such as word frequency occurrences in the query, the retrieved documents and the Web as a whole (Stanley, 1997), location of words on the page and the number of links to the page.

The World Wide Web, with its intra- and interdocument hypertext links, provides a navigational tool that facilitates database browsing in contrast to database searching. Boolean-driven retrieval systems are not hospitable to browsing. They are designed to divide a database into two parts – one that contains those records matching the search statement, and the other which contains all the

other non-matching records in the database. Yet many users do not begin with such a clear view of their information requirements that they can formulate a sharp search statement. For these users it can prove very productive to browse the database (or a part of it), looking for potentially interesting material. Hypertext allows this to take place. The danger in such systems, as exemplified by the Web, is that the browser becomes disorientated – lost, so to speak – in a tangle of linked sites. Navigational aids such as the back, forward, bookmark and search history facilities on a Web browser are intended to minimize such problems.

Traditional online systems, unlike CD-ROMs or the Web, cannot provide multimedia graphics, still images, video, animation, or sound clips. Such graphic capabilities are invaluable in many information fields.

Database differences between vendors

This diversity of delivery leads to a very important observation, and one that is explored much more fully in the following chapters: database content and/or structure can differ from one medium to another, or one vendor to another. It is therefore important that the user be aware of the different configurations of the same database title that might be encountered in the market.

Tenopir and Hover (1993) have established seven criteria by which to compare differences between the same database mounted by different vendors:

- The first criterion is update frequency. INSPEC, for example, was then updated weekly on Orbit but bi-weekly on DataStar; PsycInfo was updated quarterly on the SilverPlatter CD-ROM but monthly by several online vendors.
- Second, database time coverage can vary. CAB ABSTRACTS was then available from DIALOG since 1972 but on DataStar only since 1984.
- Third, pricing formulae for the same database can vary greatly: connect time, flat fee charge, payment per retrieved record, and so on.
- Fourth, a database may be searchable as one big file or as several files. MEDLINE, for example, was then available on DIALOG as one huge file (and therefore easier to search in its entirety), while on NLM it was broken down by date into several separate files (making it therefore easier to search for, say, records only appearing in the most recent few years).
- Fifth, database content can vary. For example, although the CHEMICAL ABSTRACTS database is available on several online systems, only the version on STN includes abstracts as well as the bibliographical citation. Dissertations are to be found in the online versions of PsycInfo, but not on the PsycLit CD-ROM equivalent. A full-text omnibus news database may include items from different source publications in its various versions; if a vendor already offers a newspaper title as a separate database, then records from that paper may not be added to the omnibus version carried by that vendor.

- Sixth, value-added support features like online thesauri, document delivery, cross-file searching and so on, will not be available from all vendors.
- Seventh, records in the same database may be structured and indexed differently by various vendors.

On the Web, the search engines index pages according to different principles as well as offering different search capabilities. For example, ALTAVISTA, EXCITE and NORTHERNLIGHT attempt to index every page on a site while INFOSEEK, LYCOS and WEBCRAWLER only index sample pages. Several offer truncation, but WEBCRAWLER and LYCOS, for example, do not. Sullivan (1998) has compiled useful comparative data for seven major search engines, and Hock (1999) offers tips on searching eight major Web engines.

Controlled versus natural-language searching

Many databases continue to offer the information seeker a choice between conducting a search using the words found in the records themselves – titles, abstracts or complete texts – or using index terms that describe the content of the record while not necessarily using the natural-language words found within it. In most cases such index terms have been assigned by a human indexer and have been chosen from a list of controlled terms representing the subject area of the particular database. Web sites (other than those few that contain organized databases which utilize controlled vocabulary) do not include controlled index terms, although in some instances the site creator may have added metadata at the beginning of the document, which can include uncontrolled keywords that seek to encapsulate its content. As this is not uniformly or exclusively used by search engines for indexing purposes it is currently of questionable benefit. Furthermore, an unscrupulous site developer can deliberately add keywords of little or no relevance to the site's content just so as to attract users (a practice termed 'spamming'). The absence of indexing on most Web sites is considered by many information professionals, at least, as a grave weakness and one reason why large numbers of irrelevant documents are so often retrieved in a Web search.

When a database does offer the choice between a search on the natural-language terms found within the records themselves, and a search limited only to the index terms assigned to those records (often called descriptors), which option should the searcher choose? Natural-language searching offers the opportunity to search for the actual words and phrases employed by the author. It also provides more words on which to search, as typically a record will not be assigned more than a handful of descriptors whereas the words in a full-text article may number in the thousands. In subject areas where the vocabulary is volatile, controlled vocabulary cannot keep pace with change; only by searching on natural language can the latest terminology be applied.

From the database producer's point of view, indexing is an expensive proposition. The benefits therefore must be considerable to justify the extra work

Title	Journal	Issue	London (Eng)	Fires	Subways	Fires-Great Britain	Great Britain
UK police doubt arson caused fatal subway fire	Calgary Herald	Nov 21,1987	■	■	■		
Reports say (London) subway fire warnings were ignored	Calgary Herald	Nov 22,1987	■	■	■		
Full inquiry ordered into (London subway) tragedy	Calgary Herald	Nov 20,1987	■	■	■		
Fire safety on subway questioned	Globe and Mail	Nov 20,1987	■	■	■		
Swift spread of flames in station baffles London subway officals	Globe and Mail	Nov 20,1987	■	■	■		
London subway bans smoking, tobacco ads	Globe and Mail	Nov 25,1987	■	■	■		
Raging fire claims 32 in (London) subway	Calgary Herald	Nov 19,1987	■	■	■		
Police don't suspect arson in subway fire	Montreal Gazette	Nov 21,1987	■	■	■		
'Coctail' of gases suspected in subway fire	Globe and Mail	Nov 21,1937	■	■	■		
35 killed, 80 hurt in UK subway fire	Globe and Mail	Nov 19,1987	■		■		
Killer fire (in London subway) sparks ban on smoking	Calgary Herald	Nov 25,1987			■		
Subway inferno kills 32:'catastrophe' strikes London underground	Halifax Chronicle Herald	Nov 19,1987	■				
(London) subway fire inquiry announced	Winnipeg Free Press	Nov 20,1987	■	■	■		
Britain to hold public inquiry into subway fire that killed 30	Montreal Gazette	Nov 20,1987	■	■	■		
32 dead in London subway blaze	Montreal Gazette	Nov 19,1987	■	■	■		
Escalator carried commuters into inferno of flames, smoke	Toronto Star	Nov 19,1987	■	■	■		
UK fire chief says blaze began on subway escalator	Toronto Star	Nov 20,1987	■	■	■		
Escalator problems cited in subway fire	Toronto Star	Nov 21,1987	■	■	■		
Flaming horror in London subway (King's Cross)	Macleans	Nov 30,1987			■	■	■
An inferno in the London Underground	Newsweek	Nov 30,1987			■	■	■
Escalator to an inferno: panic and death in London's Underground	Time	Nov 30,1987			■	■	■
Total occurrence of each descriptor			17	15	20	3	3

Source: Jacsó (1992). Reproduced by permission.

Figure 1.4   Failure of controlled language

involved in preparing the database. The greatest advantage offered by indexing is the control that it imposes over the inconsistencies and redundancies in natural language. The synonyms (and near synonyms) so frequently encountered in all languages can be represented by just one index term, thus removing the searcher's need to enter all possible synonyms in order to be certain that the actual word used by the author has been entered. A few carefully chosen controlled terms can also summarize the main subjects dealt with in a record, thereby allowing the searcher to find records that are central to a particular subject rather than minor referents (the words chosen for the title may also accomplish this objective, but titles, especially in the humanities and social sciences, are not always descriptive of the

actual content). A list of controlled terms may also be very helpful to a searcher who is unfamiliar with the subject area of the database and finds it difficult to formulate a search unaided using natural-language terms. Controlled terms from a well-organized thesaurus can help with hierarchical or generic searches. For example, someone looking for information on dogs might expect that all records dealing with individual breeds will have been indexed with the generically higher term 'dogs'; a natural-language search, to be comprehensive, might have to include the names of individual breeds in case the authors themselves only mentioned the breed and not the species. Of course, if the searcher opts for controlled language then confidence is placed in the reliability and the consistency of the database's indexers. It is not an easy task to take an article and determine which, say, ten terms from a controlled list should be chosen to represent the subject matter of the article. Figure 1.4 shows that controlled languages are still at the mercy of the human indexer.

Interfaces

Most of the interfaces discussed in the following chapters rely on graphical devices: windows, pull-down and pop-up menus, buttons and icons. A dwindling number of CD-ROMs and OPACs, as well as traditional online systems, rely on non-graphical, DOS-based interfaces. Shneiderman (1998) lists several benefits for the user of graphical interfaces:

- control over the system
- ease of learning the system
- enjoyment
- encouragement to explore system features.

At the same time, unless designed with some care, graphic devices can seem little more than gimmicks, and their initial novelty value can soon dissipate. Furthermore, users unfamiliar with graphical interfaces can find them confusing and forbidding. It is all too easy to produce cluttered and confusing screens containing too much visual information. Careful screen layout, restrained use of colour and consistency in the application of graphical devices will all help to produce an effective, rather than a baffling, interface. Many guides to sound interface design are available (see, for example, Galitz, 1997; Mandel, 1997; Shneiderman, 1998), but unfortunately, as in other areas of human endeavour, it is often easier to propose, than to follow, guidelines. Head (1997) provides a useful discussion of graphical interfaces specifically directed at online information services.

Search evaluation

The *Manual of Online Search Strategies* provides guidance on how to conduct most effectively searches for information in a variety of subject areas. It is therefore important to discuss how a successful search might be measured, so as to differentiate it from an unsuccessful search. How can search performance be evaluated?

The first large-scale tests of information retrieval systems began in the late 1950s at the Cranfield College of Aeronautics in England. The Cranfield Projects employed two measurements for an information retrieval system: recall and precision. The assumption behind these measures is that the average user wants to retrieve large amounts of relevant materials (producing high recall) while simultaneously rejecting a large proportion of irrelevant materials (producing high precision).

Recall and precision

Recall is a measure of effectiveness in retrieving all the sought information in a database – that is, in search comprehensiveness. A search would achieve perfect recall if every single record that should be found in relation to a specific query is indeed traced. It is normally expressed in proportional terms. The recall ratio in any search can theoretically be improved by finding more and more records; in fact, 100 per cent recall can always be achieved by retrieving every single record in the database, including all the irrelevant alongside the relevant ones, although this defeats the purpose of a retrieval system. Clearly, a parallel measure is required to work alongside recall, which will take account of the false hits produced. This measure is called precision. It assesses the accuracy of a search – that is, the extent to which the search finds only those records that should be found, leaving aside all records that are not wanted. A search would achieve perfect precision if every single record retrieved in relation to a specific query were indeed relevant to that query. Precision, like recall, is normally expressed in proportional terms.

The Cranfield tests found an inverse relationship between recall and precision. As attempts are made to increase one, the other tends to decline: higher recall can only be achieved at the expense of a reduction in precision. As a strategy is implemented to retrieve more and more relevant records there is a tendency also to retrieve growing numbers of irrelevant records; recall is improved but precision worsened. As a strategy is implemented to eliminate irrelevant records there is a tendency also to eliminate relevant ones; precision is improved but recall worsened. There is a common sense logic to this inverse relationship that has been demonstrated in many, but not all, evaluation tests.

Criticisms of recall and precision measures

Despite the widespread use of recall and precision as measures of search effectiveness, a number of criticisms can be made. First, recall and precision offer an incomplete evaluation of information retrieval, at least from the average searcher's point of view. Searchers may want to maximize both recall and precision, but what about other factors such as the expense involved in completing the search, the amount of time taken, and the ease of conducting it? A retrieval system might give impressive recall and precision ratios yet be costly, slow and frustrating.

Secondly, recall depends on the assumption that a user wishes to find as many relevant records as possible. In practice, users may not always want a search that finds everything, but instead opt for a search that retrieves just a few highly relevant items. In such cases, precision alone is the only measure of retrieval effectiveness. Many Web searchers, for example, are likely to find themselves in this situation; high recall in many cases will simply overwhelm the searcher.·

Thirdly, to measure recall it is necessary to know the total number of relevant records in the database, retrieved and not retrieved. But how can the number of relevant non-retrieved records in the database be established? In the case of the very small test databases sometimes used for evaluation experiments, it is feasible to examine all the documents and thus to determine which are, and which are not, relevant to any particular search query. Considerable doubt has been expressed, however, about the validity of extrapolating search results from these test databases to the much larger databases typically encountered in real searches. The constantly changing content on the Web, as well as the vast size of the 'database', make recall measurement especially problematic. Clarke and Willett (1997) have proposed, however, a methodology for measuring recall, as well as precision, in order to evaluate the effectiveness of search engines.

The most serious criticisms of precision and recall measures, however, concern the reliability of the crucial concept underlying both – relevance. How is relevance to be determined, and by whom? Experimentally, relevance judgements have typically been based on a match between the subject content – the 'aboutness' – of a retrieved record and the initial query that stimulated the search. In the Cranfield tests, for example, the subject content of the query and the subject content of the records were compared by subject experts to decide whether a retrieved record was relevant or not. Many searches are now conducted by end users and not, as in earlier days, by information intermediaries. This means that the person judging the retrieved results as they are displayed is the ultimate user of the information, who may bring various subjective elements into play. A bibliographic record, for example, might be judged relevant on the basis of its author, the series in which it appears, its recency, local availability and so on, as well as on its subject content. On the other hand, would a retrieved document that the seeker has already read be considered relevant, even if directly related to the subject? Lancaster and Warner (1993) prefer to distinguish between relevance and a related concept – pertinence. They define pertinence as the relationship between

a document and a request, based on the subjective decision of the person with the information need. They argue that pertinence decisions are essential to the evaluation of operating (rather than trial) information retrieval systems serving real users who have real information needs. Harter (1992) proposes the term 'psychological relevance' for records that suggest new cognitive connections, fruitful analogies, insightful metaphors or an increase/decrease in the strength of a belief. He argues that records about a topic may in fact prove less important to the user than relevant records which are not on the topic but that allow new intellectual connections to be made or cause other cognitive changes in the user. Furthermore, he believes that such a view of psychological relevance is inconsistent with the notion and utility of fixed relevance judgements and with traditional retrieval testing as exemplified by the Cranfield tests and their successors.

Nevertheless, most information retrieval experts do agree that subject aboutness is still the principal criterion used in judging relevance. A search system can only be judged in terms of whether it is able to match the user's information need as expressed in the search strategy with the stored data. The additional facility to screen out from the retrieved records those that the user has already read may well be a highly valuable feature, but the failure of a system to undertake this extra step cannot reasonably be invoked to judge the performance of the system at *retrieving* relevant information.

Another problem with relevancy is that judgements about one record may be influenced by other records that have already been examined. After examining nine totally irrelevant documents, a tenth one might be considered relevant, but had this tenth record been viewed after seeing nine highly relevant ones it might have been judged irrelevant. This emphasizes, of course, the binary nature of relevancy judgements for recall and precision purposes: there is no place for the fairly relevant, or the marginally relevant, or even the extremely relevant. A record is accepted as relevant or rejected as irrelevant.

Spink and Greisdorf (1997) argue that researchers should question the assumption that users always even need the most highly relevant items. At the outset of an information-seeking process a user's information problem is often ill-defined. The retrieved items considered highly relevant may well provide users with what they already know: as they are likely to equate strongly with the current state of the user's information problems, they may only reinforce the current state of the information problem. Items that are only 'partially relevant' may then play a greater role in shifting the user's thinking about the information problem, providing new information that leads the user in new directions towards the ultimate resolution of the information problem.

Unfortunately, the critics of recall/precision measures are unable to proffer any alternative quantitative evaluation technique. Yet everyone does agree that the ability to evaluate information retrieval is crucial. All this suggests that recall and precision ratios as reported in experimental studies should be treated as relative, rather than absolute, indicators. The measures of recall and precision based on estimates of relevance remain valid evaluation parameters even if their precise

measurement in experimental studies is problematic. In evaluating strategies and reacting to preliminary results during an interactive search, for example, the concepts of recall and precision are extremely useful to help the searcher decide on strategy adjustments. A small number of hits may suggest a need to broaden the search to improve recall, even if this adversely affects precision. A search with higher recall but a large percentage of irrelevant records is a prime target for strategy adjustments to improve precision, even if at the expense of lower recall.

Making judgements during a search on the relevance of intermediate results and then using these judgements to revise the search strategy is termed 'relevance feedback'. Some information retrieval systems do not simply rely on the searcher to initiate such feedback. The system itself may automatically search, for example, to find more records that share index terms with records already retrieved and judged relevant by the user.

Database evaluation

No matter how sophisticated is the information seeker or powerful the retrieval system, the resulting information from a search will ultimately depend on the database quality: the accuracy, completeness, authority and currency of its information, and the reliability of its indexing. Unfortunately, database quality cannot be taken for granted. Information stored in electronic format is inherently no more nor less reliable or accurate than other kinds of information. A few years ago, when much electronic information had been transcribed at the keyboard from hard copy originals, many typographical errors were detected in all kinds of databases. These errors not only affected data use but also data retrieval; only by similarly misspelling the term in the search would the record containing the misspelled word be found. More data are now generated at the outset electronically, and scanning equipment is more reliable, but errors are still to be found in databases.

Technically, electronic data can be updated more easily and quickly than, say, printed information. Many databases are updated monthly, weekly, daily or even in real time. Nevertheless, it should not be assumed that electronic information is always current. A number of online encyclopaedias, for example, contain data long since superseded even by print sources. Typically, online or Web-based sources are updated more frequently than CD-ROM or print equivalents (if these also exist), but occasionally technical problems have reversed this dictum.

The Internet, in particular, has highlighted the problems of data reliability. There is no umbrella organization to ensure data accuracy, currency or consistency, or to vouch for the authority of the data. Now that anyone can create a Web site there is no longer an established publishing process to ensure some kind of quality control through, for example, market pressures or academic refereeing. It can be difficult to assess the validity of much data available on the Web, although some search engines claim to exercise judgement when deciding whether
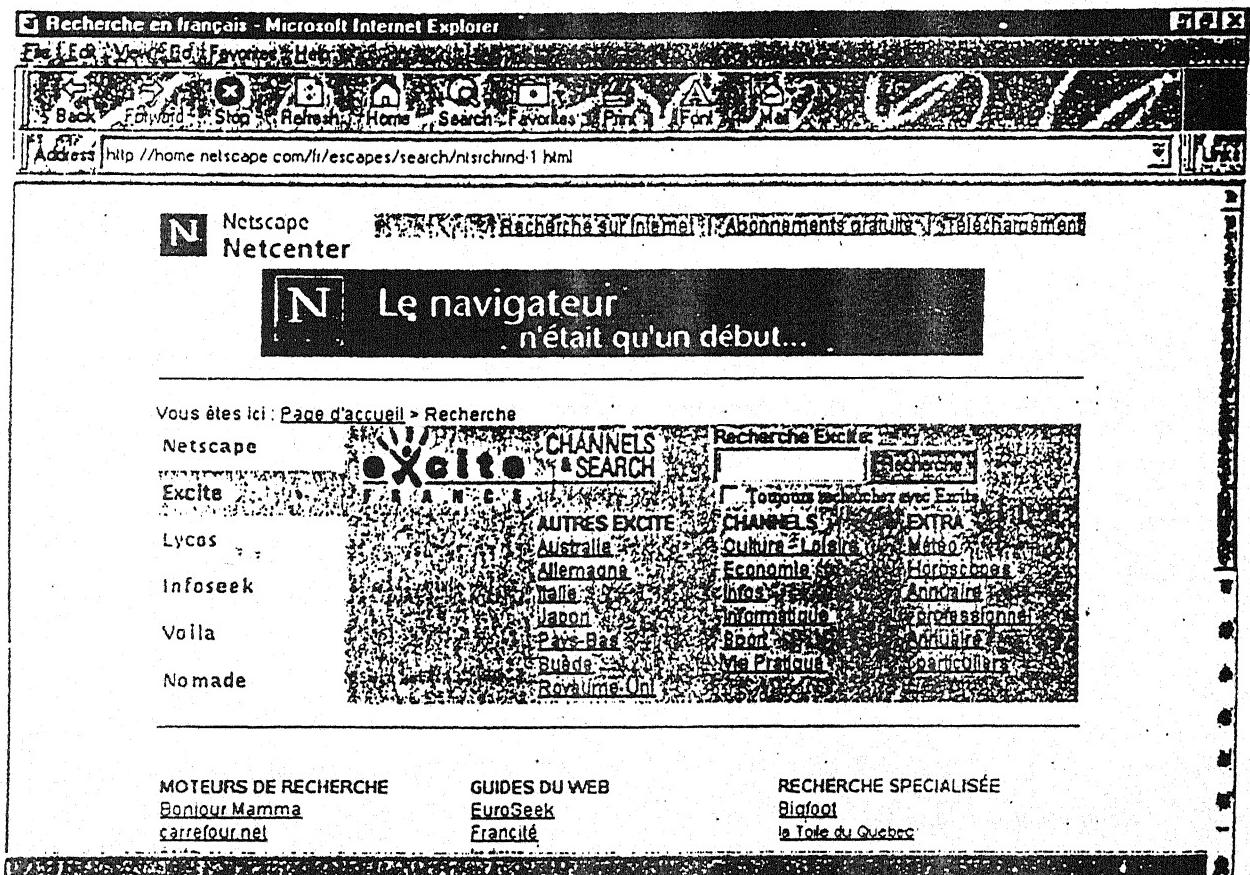
Figure 1.5   French-language interface to EXCITE

or not to provide access to sites from their hierarchical indexes, and some provide evaluation scores based on an assessment of 'quality'. The ephemeral nature of much Web material also means that what can be found today may have vanished or been transformed by tomorrow. In Volume I, Chapter 5 on the biosciences, for example, Frank Kellerman cites the example of the GENOME database from Johns Hopkins School of Medicine which had a short-lived life on the Web before being withdrawn. The ephemeral nature of much Web content is now posing very considerable bibliographic problems.

Although the Web is dominated by English-language information, pages in other languages are now being added proportionately faster than the English-language sites. Many problems remain in accessing, on the Web, information in a language unknown to the searcher (Large and Moukdad, 2000) but the reality of a truly worldwide service, free from language barriers, is getting a little closer. The provision, by many Web search engines, of interfaces in languages other than English, for example, is a welcome sign. Figure 1.5 shows the French-language interface from EXCITE. It is now possible to input search terms in other languages and to confine the resulting search either to sites in that language or to sites originating in a country using that language. ALTAVISTA goes one step further by allowing a search term in one language to be translated automatically into the corresponding term in a second language (although currently only between
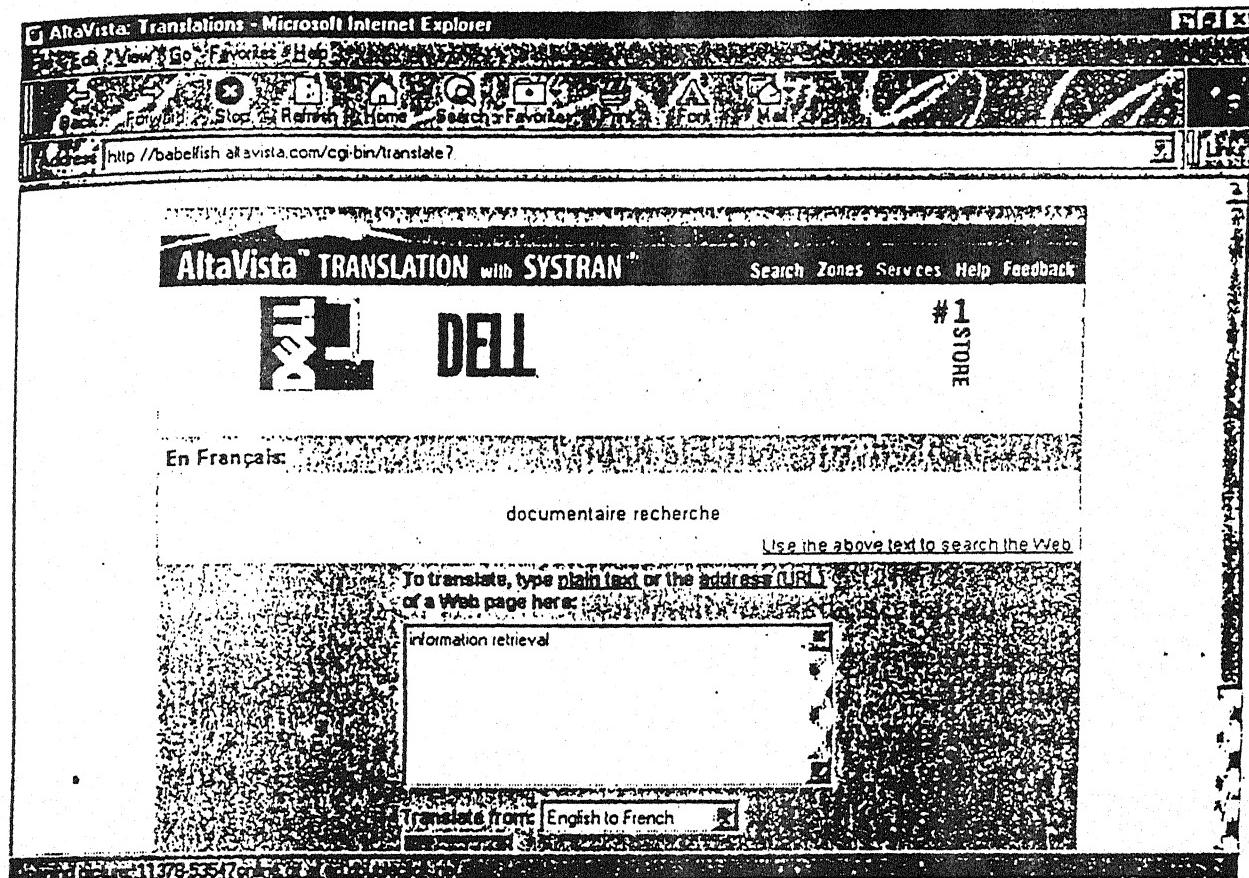
Figure 1.6   SYSTRAN translation software on ALTAVISTA

English, on the one hand, and French, German, Italian, Portuguese or Spanish on the other) prior to a Web search. In Figure 1.6 the term 'information retrieval' has been translated into French by SYSTRAN, the machine translation software employed by ALTAVISTA: the correct French phrase requires an inversion of word order from the English original: 'recherche documentaire' rather than 'documentaire recherche' as SYSTRAN has it (although this does not matter in a Web search unless the two words are to be searched together as a phrase).

The recurrent issues concerning pornography and hate literature further complicate discussions concerning Internet content. This is likely to prove especially controversial where certain user groups, and especially children, are involved. Some libraries have experimented with various proprietary site-blocking software, but the results generally prove unsatisfactory, failing to block some material deemed unacceptable while excluding other material to which exception is not taken.

Evaluation criteria

Evaluation criteria for databases have been proposed by a number of authors, and there is a large measure of agreement about these. One influential evaluation

checklist was formulated in 1990 by the Southern California Online User Group (Basch, 1990):

- Consistency – does the database maintain consistency in coverage, currency and so on? If it is one of a family of databases, how consistent are these products in interface design, update policy and such like?
- Coverage/scope – does the coverage/scope match the stated aims of the database; is coverage comprehensive or selective?
- Error rate/accuracy – how accurate is the information?
- Output – what kind of output formats are available?
- Customer support and training – is initial or ongoing training provided? Is a help desk available during suitable hours?
- Accessibility/ease of use – How user-friendly is the interface? Does it have different facilities for novice and experienced searchers? How good are the error messages? Are they context-sensitive?
- Timeliness – is the database updated as frequently as it claims, and as the data warrant?
- Documentation – is online and/or printed documentation clear, comprehensive, current and well-organized?
- Value to cost ratio – finally, taking into account the above features, does the database give good value for money?

Anagnostelis and Cooke (1997) propose somewhat more detailed evaluation criteria to be applied to Web-based databases – in this case, specifically for comparison of various MEDLINE database services on the Web:

- authority of the service provider as well as the database
- content – coverage and currency
- retrieval mechanism – general search features, free-text searching, natural-language queries, thesaurus searching, command searching, display and output
- ease of interface use
- unique features
- help and user support.

Similar evaluation criteria can be found in Tenopir and Hover (1993) who specifically discuss comparison of the same database available on different systems, and from the Organising Medical Networked Information (OMNI) Consortium (1997), which is involved, among other tasks, with the evaluation of MEDLINE Services on the Web. The SOCIAL SCIENCE INFORMATION GATEWAY has prepared a detailed list of criteria used by various eLib Gateway projects (including OMNI as well as SOSIG itself), while Bartelstein and Zald (1997), and Tillman (1997) also provide valuable insights. A longer discussion can be found in Cooke (1999).

47

Web search: how the Web has changed information retrieval

Terrence A. Brooks
Information School
University of Washington
Seattle, Washington, USA

Abstract

Topical metadata have been used to indicate the subject of Web pages. They have been simultaneously hailed as building blocks of the semantic Web and derogated as spam. At this time major Web browsers avoid harvesting topical metadata. This paper suggests that the significance of the topical metadata controversy depends on the technological appropriateness of adding them to Web pages. This paper surveys Web technology with an eye on assessing the appropriateness of Web pages as hosts for topical metadata. The survey reveals Web pages to be both transient and volatile: poor hosts of topical metadata. The closed Web is considered to be a more supportive environment for the use of topical metadata. The closed Web is built on communities of trust where the structure and meaning of Web pages can be anticipated. The vast majority of Web pages, however, exist in the open Web, an environment that challenges the application of legacy information retrieval concepts and methods.

Introduction

Search is a compelling human activity that has extended from the Library at Alexandria to the World Wide Web. The Web has introduced millions of people to search. The information retrieval (IR) community stands ready (Bates, July 2002) to suggest helpful strategies for finding information on the Web. One classic IR strategy - indexing Web pages with topical metadata - has already been tried, but the results are disappointing. Apparently, relying on Web authors to garnish their Web pages with valid topical metadata runs afoul of human nature:

48

- Sullivan (2002c) reports that the meta *keywords* tag, an HTML element designed for adding descriptors to Web pages, is regarded as untrustworthy and avoided by all major search engines.
- A FAQ at the Dublin Core site explains that well-known 'all the Web' search engines 'tend to avoid using the information found in meta elements' for fear it is spam (FAQ).

Applying topical metadata to Web pages provokes a controversy that pits partisans who envisage a semantic Web featuring topic maps and ontologies of shared meanings (Berners-Lee, Hendler & Lassila, May 2001) versus detractors who disdain topical metadata as 'metacrap' (Doctorow, August 26, 2001) and warn us of a Web of deception (Mintz, 2002). The significance of the controversy, however, awaits the examination of a more fundamental issue: does it make *technological* sense to add topical metadata to Web pages?

If the Web is a big, distributed document database and Web pages are composed in HTML (i.e., 'the document in my browser goes from <html> down to </html>'), the answer is 'yes.' In this case, it makes technological sense for Web authors to add topical metadata to Web pages, just as an indexer might add descriptors to a document in a database. An affirmative answer validates the topical metadata debate. If, however, the Web is not a big document database, but is instead a network of rapidly changing presentations, the answer is 'no.' In this view HTML is primarily a presentation technology, and most Web pages are transitory and volatile presentations governed by the whims of viewer taste and the contingencies of viewer technology. A negative answer signals that debating the value of topical metadata is premature until it can be shown

that they are technologically appropriate additions to Web pages.

Lurking behind the topical metadata controversy is our unsteady application of the concept of 'document' to Web content and presentation. We inherit our notion of document from vertical-file systems and document databases, technological environments not known for schisms between content and presentation. Viewed from the document-database tradition, indexing Web pages appears to be a simple extension of current practice to a new, digital form of document. Viewed from the HTML tradition, however, indexing Web pages confuses presentation for content. Topical metadata are intended to index information content, not arbitrary or personalized views of content, and the majority of Web pages are arbitrary presentations contingent on Web browsers, security settings, scripts, plug-ins, cookies, style sheets and so on.

Considering the appropriateness of the document metaphor for the Web has fundamental consequences for the application of IR's extensive body of theory and practice. Controversies about topical metadata aside, recognizing the familiar IR notion of 'document' on the Web would suggest that Web searchers are retrieving information, and that we can apply IR concepts and methods to help Web searchers. In this case, the topical metadata controversy gains significance. Realizing that the document metaphor does not map to the Web, however, heralds a paradigm shift. Perhaps Web searchers are not retrieving information, but

doing something else. 'Web search' is used in this essay to name the activity of discovering, not retrieving, information on the Web.

IR and the 'document' metaphor

The technological legacy of search

The foundation of search in the last century has been the storage and retrieval of paper based on some form of labeling. Yates (2000) describes vertical filing that made information accessible by using labeled files to hold one or more papers:

Vertical filing, first presented to the business community at the 1893 Chicago World's Fair (where it won a gold medal), became the accepted solution to the problem of storage and retrieval of paper documents....The techniques and equipment that facilitated storage and retrieval of documents and data, including card and paper files and short- and long-term storage facilities, were key to making information accessible and thus potentially useful to managers. (Yates. 2000: 118 -120)

The application of computer databases to search by mid-20th century extended the vertical file paradigm of storage and retrieval. A computer database is a storage device resembling a vertical file just as a database record is a unit of storage resembling a piece of paper. The more abstract term 'document' addressed any inexactitude in the equivalence of 'database record = piece of paper." Computer databases were seen as storing and retrieving documents, which were considered to be objects carrying information:

51

- 'Information retrieval is best understood if one remembers that the information being processed consists of documents.' (Salton & McGill, 1988, p. 7)
- 'With the appearance of writing, the document also appeared which we shall define as a material carrier with information fixed on it.' (Frants et al., 1997: 46)
- 'Document: a unit of retrieval. It might be a paragraph, a section, a chapter, a Web page, an article, or a whole book.' (Baeza-Yates & Ribeiro-Neto, 1999: 440)

Digitizing documents greatly boosted the systematic study of IR. Texts could be parsed to identify and evaluate words, thereby perhaps discovering meaning. Facilitating assumptions about the nature of documents and authorial strategies were advanced. For example, Luhn (1959: 160) suggested that 'the frequency of word occurrence in an article furnishes a useful measurement of word significance.' In the following extract Salton and McGill (1988) suggest where subject topical terms are located in documents, and how text can be processed to find these terms:

The first and most obvious place where appropriate content identifiers might be found is the text of the documents themselves, or the text of document titles and abstracts.... Such a process must start with the identification of all the individual words that constitute the documents.... Following the identification of the words occurring in the document texts, or abstracts, the high-frequency function words need to be eliminated... It is useful first to remove word suffixes (and possibly also prefixes), thereby reducing the original words to word stem form. (Salton & McGill, 1988: 59, 71).

The document-database search technology sketched above maps easily to the Web and suggests that searching on the Web is an extension of IR:

- Vast numbers of documents are available on the Web (e.g., 'the Web is a big database.')

52

- Viewing the source of a Web presentation reveals a structured document (e.g.: 'the document goes from <html> down to </html>.')
- Google seems to index Web pages (e.g., 'Google is a big index made up of words found in Web pages.')

The legacy social context of search

We inherit, as well, an elaborate social context of search that has been applied to the Web. Librarianship was the source of powerful social conventions of search even before the introduction of the technology of vertical files. For example, Charles A. Cutter suggested rules for listing bibliographic items in library catalogues as early as 1876. Bibliographic standardization, expressed in the Anglo-American Cataloguing Code, was a powerful idea that promoted the view that the world could cooperate in describing bibliographic objects. An equally impressive international uniformity was created by the wide acceptance of classification schemes, such as the Dewey Decimal Classification (DDC):

Other influences are equally enduring but more invisible, and some are especially powerful because they have come to be accepted as 'natural.' For example, the perspectives Dewey cemented into his hierarchical classification system have helped create in the minds of millions of people throughout the world who have DDC-arranged collections a perception of knowledge organization that had by the beginning of the twentieth century evolved a powerful momentum. (Wiegand, 1996: 371)

The application of computer databases by mid-20th century spurred many information communities to establish or promote social conventions for their information. For example, the Education Resources Information Center (ERIC), 'the world's largest source of education information' (Houston, 2001: xiv), represents a community

53

effort to structure and index the literature of education. At the height of the database era in the late 1980s, vendors such as the Dialog Corporation offered access to hundreds of databases like ERIC, each presenting one or more literatures structured and indexed. This social cooperation and technological conformity fostered the impression that, at least in regards to certain subject areas, the experts had their information under control.

The social context of document-database search sketched above maps easily to the Web and suggests a benign, socially cooperative information environment:

- Web authors will add topical metadata to their Web pages (e.g., 'I index my Web pages with keywords and Dublin Core metadata so people will find them on the Web.')
- Everyone will use topical metadata (e.g., 'The semantic Web will be constructed by millions of Web authors indexing their Web pages.')
- Web crawlers, like Google, will harvest topical metadata (e.g., 'Google has indexed my topical metadata and now my Web pages are available for retrieval.')

We are now just learning that the Web has a different social dynamic. The Web is not a benign, socially cooperative environment, but an aggressive, competitive arena where authors seek to promote their Web content, even by abusing topical metadata. As a result, Web crawlers must act in self defense and regard all keywords and topical metadata as spam.

Debating whether topical metadata are spam or an essential step towards the construction of the semantic Web assumes that they are technologically appropriate additions to Web

pages. To what extent are Web pages analogues of the legacy IR document-container of information?

The Web and the 'document' metaphor

A Web page is a 'snapshot'

Documents added to the ERIC database thirty years ago are still retrievable. There is every expectation that they can be retrieved next year. This expectation provides a rough definition of what it means to retrieve information – finding the same document time and again. The metaphor used in the working draft on the *Architectural Principles of the Web* (Jacobs, August 30, 2002), however, does not suggest retrieving the same thing time and again. Interacting with a Web resource gives one a snapshot:

> There may be several ways to interact with a resource. One of the most important operations for the Web is to retrieve a representation of a resource (such as with HTTP GET), which means to retrieve a snapshot of a state of the resource. (Jacobs, August 30, 2002, section 2.2.2)

Web resources are characterized as evolving, not static, resources. They are more like loose-leaf binder services than time-invariant database records:

> An integrating resource is a bibliographic resource that is added to or changed by means of updates that do not remain discrete and are integrated into the whole. Examples of integrating resources include updating loose-leafs and updating Web sites. (Task group on implementation of integrating resources, 2001)

If Web pages are snapshots then the critical question is rate of update. Some ERIC records are thirty years old; the oldest HTML pages date from about ten years ago, but most Web content is much more volatile:

- Brewington and Cybenko (2000) observed that half of all Web pages are no more than 100 days old, while only abut 25% are older than one year.
- Cho and Garcia-Molina (2000) found 40% of Web pages in the *.com* domain change everyday. The half-life of Web pages in the *.gov* and *.edu* is four months.
- Koehler (1999) found the half-life of Web content is two years.
- Spinellis (2003) found the half-life of URLs is four years.
- Markwell and Brooks (April 15, 2002) found the half-life of science education URLs to be fifty-five months.
- Cockburn and McKenzie (2001) found that the half-life of bookmarks to be two months.

Content churn and rapid birth and death cycles distinguish Web pages from the legacy IR document-container of information. Philosophers can address the issue of repeated refreshing of the 'same' Web page that presents 'different' content each time, as to whether this is the 'same' Web page or 'different' Web pages. Whatever grist falls from the philosophical mill, it is clear that Salton and McGill didn't consider database documents to be snapshots.

Web pages are cultural artifacts

Web content is only available through the mediation of a presentation device, such as a Web browser. Complicating Web presentation are security settings, different computer monitors, safe and unsafe Web colors, plug-ins, cookies, scripts, and so on. In fact, Web authors expend enormous amounts of time and energy engineering a consistent presentation across platforms.

The representations of a resource may vary as a function of factors including time, the identity of the agent accessing the resource, data submitted to the resource when interacting with it, and changes external to the resource.' (Jacobs, August 30, 2002, section 2.2.5)

Figure 1 illustrates the process of converting HTML to a browser display for the Mozilla layout engine (Waterson, June 10, 2002).
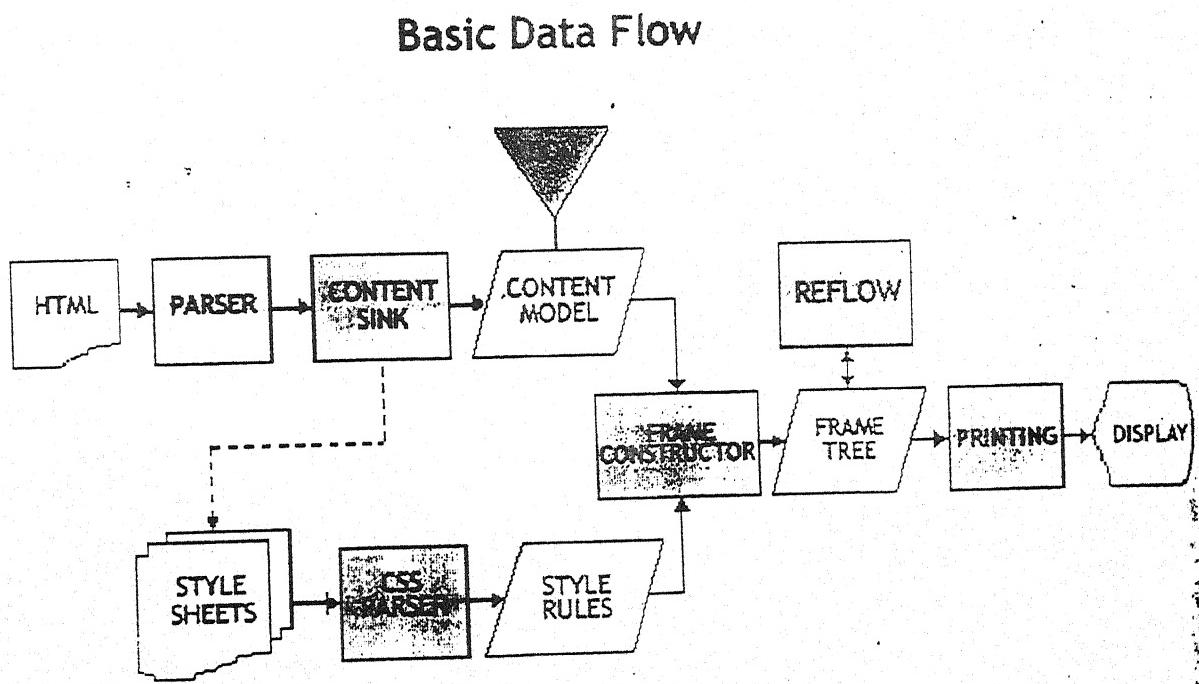
Basic Data Flow

Figure 1: Basic data flow in Mozilla layout engine (Waterson, June 10, 2002)

This diagram illustrates that HTML code is parsed and deconstructed into a hierarchical content model. Style sheets that reference content elements are also parsed. A frame constructor mixes content with style rules into a hierarchy of content frames. Nested content frames are painted to create the presentation in a Web browser. Figure 1 implies that different HTML parsing rules, style sheet applications, frame construction algorithms, and so on, would produce a different presentation.

Figure 1 also implies that assembling Web content to look like a printed document is not a technical necessity, but a

cultural convention. If your Web browser presents you with something that looks like a printed page, it is because the engineers of Web browsers are obeying the cultural expectations of the majority of their users; that is, information should resemble the familiar printed page. The mutability of Web presentation is not deplored, but actually trumpeted as an advantage in delivering customized appearance. In short, Web content can be made to look like your *favorite printed page*:

Cookies serve to give Web browsers a 'memory', so that they can use data that were input on one page in another, or so they can recall user preferences or other state variables when they user leaves a page and returns. (Flanagan, 1997, p.231)

Finding the Web page in your browser located between <HTML> and </HTML> tags reflects how your browser constructed the byte stream from the source server machine, but says nothing about how the content was structured on the source server machine. The source content could be distributed among a number of databases, XML documents, scripts, files and so on. XSLT style sheets can assemble Web site 'skins' from databases and XML sources with equal ease (Pierson, March 2003)

During the early years of the Web, most Web pages were constructed in HTML and many handcrafted Web pages are still written this way. Efficiencies of scale, however, have forced large producers of Web content to automate Web page production:

- Turau (1999) speculated that 75% of Web pages are generated from databases.

- Bergman (2001) describes the 'deep' Web as 400 to 500 times larger than the 'surface' Web. The deep Web is composed of database generated pages.

Web pages are presentation contingencies and server programming artifacts. This schism between content and presentation distinguishes them from the legacy IR notion of document-container of information. The document in your Web browser may look like a document, but probably has no documentary origin at all.

Google is not a Web index

An index helps searchers find information in a database. It is generally true that success in finding information in a database is directly proportional to knowledge about how the database documents were indexed. Database vendors such as the Dialog Corporation are famous for running classes helping searchers understand how information is indexed. Google is a popular search tool for Web content, twice voted most outstanding search engine by the readers of Search Engine Watch. In August 2002, about 28% of Web search was done with Google (Sullivan, 2002b). 'Google gets 150 million queries a day from more than 100 countries' (Harmon, February 17, 2003). Google is famous for presenting results according to page rank, but nobody knows how Google's parsing algorithm works.

Sullivan (2002a) surmises that Google uses over 100 factors to parse Web content, which still includes 'traditional on-the-page factors.' (The algorithm of Salton and McGill focuses on these factors.) If Google were to expose its parsing algorithm, it would be immediately

59

exploited by Web authors seeking to gain advantage and visibility for their Web content. Google's economic viability depends maintaining this secret: a corporate strategy strikingly different from legacy database vendors like the Dialog Corporation. Google warns Web authors who would attempt to ferret out and exploit their parsing algorithm:

We will not comment on the individual reasons a page was removed and we do not offer an exhaustive list of practices that can cause removal. However, certain actions such as cloaking, writing text that can be seen by search engines but not by users, or setting up pages/links with the sole purpose of fooling search engines may result in permanent removal from our index. (http://www.google.com/Webmasters/2.html).... Google's complex, automated methods make human tampering with our results extremely difficult (http://www.google.com/technology/index.html).

Google does not attempt to cover the entire Web. It systematically excludes Web sites with doorway pages or splash screens, frames and pages generated 'on-the-fly' by scripts and databases. Jesdanun (October 25, 2002) reports content removed from Google to satisfy national prohibitions. Many Web pages also include objects Google finds opaque such as image files and applets. Increasing numbers of Web presentations have no text at all: 'Graphic design can be content where users experience a Web-site with little or no "text" *per se*'. (Vartanian, 2001).

Legacy indexing algorithms were open for inspection. Adding topical metadata to Web pages in the hope that Google will harvest them is betting on an unknown indexing strategy. Google will no tell you what it did with your topical metadata because being a black box is a corporate survival strategy.

Conclusion

The preceding survey of Web technology indicates that Web pages make poor hosts for topical metadata. This is not an evaluative judgment about topical metadata themselves, but merely an observation that they are misapplied to a technology characterized by churning content in arbitrary presentations parsed by unknowable algorithms. The cost and effort of adding topical metadata to an information structure is only recouped if that information structure persists in time with a predictable structure, identity and contents. An example of such a structure is the legacy IR document-container of information. Topical metadata await their more appropriate application on the Web in environments where the technical and social factors supporting the IR document-container of information can be re-created. This can be done by 'closing' the Web.

Closing the Web to do information retrieval

The legacy technical and social environments supporting IR in document databases are sketched above. It is possible to re-create this environment on the Web behind passwords in venues such as intranets, enterprise computing, and digital libraries. These applications are driven by social groups that can reach agreements on information structure and topical metadata. For example, a social group can arbitrarily decide to construct and present its information in HTML or any other presentation technology. It can also decide to use the meta keywords tag or Dublin Core

metadata with its choice of thesaurus of indexing terms and phrases.

Social agreements take precedence over technology in closed Webs where the Web is reduced to a communications venue. Predictability in structure and meaning is the fundamental facilitator permitting in-house Web crawlers to harvest topical metadata for the retrieval benefit of the local community. In a closed Web, one can build a legacy database and do IR.

If our terms of reference are a closed Web, then the topical metadata controversy recognizes topical metadata as important elements of a semantic (i.e., 'closed') Web.

eb search on the 'open' Web

Now and in the future, Google and similar tools will scan billions of Web presentations on the open Web. Everyday searchers will use Google to find information, and many will characterize their activity as retrieving information, despite disappearing Web pages, rotten links and Web content that changes on each viewing.

The open Web is a network where the cost of entry is merely access to a server machine. There are no social conventions about who can author a Web presentation or what can be presented. It is an unconstrained environment where initiatives requiring Web authors to add indexing terms and phrases or to structure their Web pages a certain way are doomed to failure, or will be exploited by the unscrupulous. In the open Web there is no guarantee that

62

Web presentations will remain or that servers will continue to function.

Many will use the Web to 'retrieve information,' but they are engaged in Web search, a process of constant discovery, not retrieval. The only way to preserve a Web presentation is to cache it, which is to take a *snapshot of a snapshot* and thereby create a new static representation of a continuously evolving process.

The open Web challenges us to ransack our IR legacy of concepts and methods to find any that can be applied. But it is possible that the open Web is so novel a technological platform that we will be forced to recognize that our IR legacy of concepts and methods has been historicized to the modern database era of the late 20th century.

References:

- Baeza-Yates, R. & Ribeiro-Neto, B. (1999) *Modern information retrieval*. Reading, MA: Addison-Wesley.
- Bates, M.J. (2002). After the dot-bomb: getting Web information retrieval right this time. *First Monday*, 7 (7) http://firstmonday.org/issues/issue7_7/bates/index.html (20 March 2003)
- Bergman, M. K. (2001) The deep Web: surfacing hidden information *The Journal of Electronic Publishing*, 7 (1) http://www.press.umich.edu/jep/07-01/bergman.html (18 January 2003)
- Berners-Lee, T., Hendler, J. & Lassila, O. (2001) The semantic Web. *Scientific American*, **284**(5), 34-+ http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2 (20 March 2003)
- Brewington, B.E. & Gybenko, G. (2000) How dynamic is the Web? In: [Proceedings of the] Ninth International World Wide Web Conference, Amsterdam May 15-19, 2000. WWW9.org. http://www9.org/w9cdrom/264/264.html (20 March 2003)
- Cho, J. & Garcia-Molina, H. (2000) The evolution of the Web and implications for an incremental crawler. In: Amr El Abbadi, Michael L.

Brodie, Sharma Chakravarthy, Umeshwar Dayal, Nabil Kamel, Gunter Schlageter, and Kyu-Young Whang, editors. *VLDB 2000, Proceedings of 26th International Conference on Very Large Data Bases, September 10-14, 2000, Cairo, Egypt.* San Francisco, CA: Morgan Kaufmann. http://rose.cs.ucla.edu/~cho/papers/cho-evol.pdf (20 March 2002)

- Cockburn, A. & McKenzie, B. (2001) 'What do Web users do? An empirical analysis of Web use' *International Journal of Human-Computer Studies*, **54**(6), 903-922

- Doctorow, C. (2001) Metacrap: putting the torch to seven straw-men of the meta-utopia. [Personal Web site] http://www.well.com/~doctorow/metacrap.htm (20 March 2003)

- FAQ: What search-engines support the Dublin Core Metadata Element Set? Dublin Core Metadata Initiative. http://dublincore.org/resources/faq/#whatsearchenginessupport (20 March 2003)

- Flanagan, D. 1997. *JavaScript: the definitive guide.* Sebastopol, CA: O'Reilly.

- Frants, V.I., Shapiro, J. & Voiskunskii, V.G. (1997) *Automated informaton retrieval: theory and methods.* San Diego, CA: Academic Press.

- Harmon, A. (2003) Google deal ties company to Weblogs. *The New York Times* Monday, February 17, C3.

- Houston, J. E. (2001) *Thesaurus of ERIC descriptors*, 14th edition. Westport, CT: Oryx Press.

- Jacobs, I. (2002). *Architectural principles of the World Wide Web. W3C working draft..* World Wide Web Consortium (W3C.org) http://www.w3.org/TR/2002/WD-Webarch-20020830/ (20 March 2003)

- Jesdanun, A. (2002) Report: sites missing from Google. News.com.au http://www.news.com.au/common/printpage/0,6093,5356761,00.html (20 March 2003)

- Koehler, W. (1999) Digital libraries and the World Wide Web sites and page persistence. *Information Research*, 4, (4) http://informationr.net/ir/4-4/paper60.html (23 March 2003)

- Luhn, H. P. (1959) The automatic creation of literature abstracts. *IBM Journal of Research and Development* 2(2), 159-165.

- Markwell, J. & Brooks, D.W. (2002) Broken links: just how rapidly do science education hyperlinks go extinct? Lincoln, NE: University of Nebraska-Lincoln. Department of Biochemistry. http://www-class.unl.edu/biochem/url/broken_links.html (20 March 2003)

- Mintz, A. (2002) Web of deception: misinformation on the Internet. Medford, NJ: Information Today.

64

- Pierson, H. (2003) <u>Site skinning: rich XML classes let users personalize their visual experience on your ASP.NET site.</u> *MSDN Magazine*, March, 87-92.
  http://msdn.microsoft.com/msdnmag/issues/03/03/SiteSkinning/default.aspx (20 March 2003)
- Salton, G. & McGill, M.J. (1988) *Introduction to modern information retrieval.* New York, NY: McGraw-Hill.
- Spinellis, D. (2003) The decay and failures of Web references. *Communications of the ACM*, 46(1), 71-77
- Sullivan, D. (2002a) <u>Google: Can the Marsha Brady of search stay sweet?</u> Search Engine Watch.
  http://searchenginewatch.com/sereport/02/09-google.html (20 March 2003)
- Sullivan, D. (2002b) <u>Nielsen//NetRatings search engine ratings.</u> Search Engine Watch.
  http://www.searchenginewatch.com/reports/netratings.html (20 March 2003)
- Sullivan, D. (2002c) <u>Death of a meta tag</u> Search Engine Watch
  http://searchenginewatch.com/sereport/02/10-meta.html (20 March 2003)
- Task Group on Implementation of Integrating Resources. (2001) *Interim report.* Washington, DC: Library of Congress.
  http://www.loc.gov/catdir/pcc/tgintegrpt.html (20 March 2003)
- Turau, V. (1999) <u>Making legacy data accessible for XMl applications.</u> [Personal Web site] http://www.informatik.fh-wiesbaden.de/~turau/ps/legacy.pdf (20 March 2003)
- Vartanian, I. (2001) *Now loading....* Coret Madera, CA: Gingko Press
- Waterson, C. (2002) <u>Introduction to layout in Mozilla.</u> Mozilla.org
  http://www.mozilla.org/newlayout/doc/gecko-overview.htm (20 March 2003)
- Wiegand, W.A. (1996) *Irrepressible reformer, A biography of Melvil Dewey.* Chicago, IL: American Library Association.
- Yates, J. (2000) Business use of information and technology during the industrial age, in *A Nation transformed by information: how information has shaped the United States from Colonial Times to the Present*, edited by A.D. Chandler & J.W. Cortada. pp. 107-136. New York, NY: Oxford University Press.

International Collaboration on Internet Subject Gateways

Emma Place

Emma Place (formerly Emma Worsfold) works at the Institute for Learning and Research Technology at the University of Bristol, UK. She works on Internet projects such as SOSIG (the Social Science Information Gateway) and DESIRE, a European project which has conducted research into Internet Services. Previously Ms Place worked for the Department of Information Studies at the University of Sheffield and for Avon County Library Service. She has degrees in Psychology and Librarianship. She is co-author of the *Internet Detective Web* tutorial and *The Information Gateways Handbook*. She has written a number of articles on Internet and electronic library topics. Ms Place may be contacted at ILRT, University of Bristol, 8-10 Berkeley Square, Bristol BS8 1HH, UK (fax: +(44-117) 9287112; e-mail: emma.place@bristol.ac.uk).

[Ms Place's paper was delivered at the 65th IFLA Council and General Conference, Bangkok, Thailand, 20-27 August 1999.]

Introduction

A number of libraries in Europe are involved in the development of Internet subject gateways - services that aim to help

users find high quality resources on Internet. Subject gateways such as SOSIG[1] (Social Science Information Gateway) have been available on the Internet for some years now, and they offer an alternative to Internet search engines such as Altavista[2] and to directories such as Yahoo.[3] Distinctively, subject gateways draw upon the skills, practices and standards of the international library community and apply these to Internet-based information. This article will suggest that librarians are ideally placed to play a major role in building Internet resource discovery services and that subject gateways offer a means to do this. It will outline some of the subject gateway initiatives in Europe and will describe the tools and technologies developed by the DESIRE[4] project to support the development of new gateways in other countries. It will also discuss how IMesh,[5] a group for gateways from around the world, aims to work on an international strategy for subject gate-

ways and on developing standards to support this.

Background

"The Web is quickly becoming the world's fastest growing repository of data."
(Tim Berners-Lee, W3C director and creator of the World Wide Web (WWW)

This is a time of upheaval for the library profession, as the Internet becomes a major medium in the information world. The Internet offers access to myriad information resources but the fact remains that it is still very hard for people to locate high quality information amid the general chaos. In the past few years the issue of resource discovery on the Internet has been the focus of much work by many different communities.

Search Engines

The Internet Search engines, such as AltaVista, and Excite, rely on automated solutions to resource discovery. They send out robots or Web crawlers to trawl the Internet and automatically index the files that they find. These indexes can then be searched by keyword and return records which contain automatically generated descriptions of the resources, usually the first few paragraphs of the resource itself. Search engines are good for finding lots of information - a search often yields thousands of resources. However, the results can be overwhelming, unmanageable, full of irrelevant references and are often too prolific to meet user needs.

Web Directories

Web directories such as Yahoo and The Open Directory[6] are, in a sense the Internet equivalent of a public library that is not staffed by librarians! They rely on human input to create directories on the WWW that

list Internet resources, with each one described briefly and classified under a subject heading. These directories aim to describe large numbers of Internet resources and include both serious and recreational sites.

The Open Directory is a remarkable project, since, in a sense, the general public are invited to build their own library on the Internet - selecting, classifying and "cataloguing" resources. The Open Directory has a *volunteer* work force of Editors (currently over 6,000 of them), who spend time adding resources and resource-descriptions to the directory (currently over 100,000!). Both Yahoo and The Open Directory aim to be the biggest Internet directories, with a high level of coverage and popular appeal as high priorities.

Internet Subject Gateways

Subject gateways offer an alternative to the Internet search engines and Web directories. What is the definition of a subject gateway? In some sense they are the Internet equivalent of an academic or special library. Subject gateways are Internet-based services designed to help users locate high quality information that is available on the Internet. They are typically, databases of detailed metadata (or catalogue) records which describe Internet resources and offer a hyperlink to the resources. Users can chose to either search the database by keyword, or to browse the resources under subject headings. Subject gateways are characterized by two key factors:

They are selective, pointing only to Internet resources that meet with quality selection criteria
They are built by subject and information specialists - often librarians.

Quality Selection Procedures

Formal quality selection criteria are used to guide collection development within the gateways. Examples of the selection policies of gateways have been collected by the DESIRE project.[7] DESIRE has also

produced an online tutorial called "Internet Detective"[8] that aims to teach the skills required to evaluate the quality of resources on the Internet and this gives some insight into the sort of work that gateway staff do in evaluating and selecting Internet resources.

Classification of Internet Resources

Classification schemes are used by gateways to set up the browsing option for users. Many gateways use traditional library classification schemes such as Dewey Decimal classification or Universal Decimal classification. A report on the use of classification schemes in Internet services has been produced by the DESIRE project, which describes this usage in more detail.[9]

Standard Metadata Formats

Standard metadata formats are used when describing an Internet resource in a database record. These formats support effective information retrieval from the databases, but also ensure that gateways can interoperate with each other and, potentially, with other databases such as library OPACS. These standards also give the option of converting and mapping one format to another, which could be important as Web metadata standards develop and change. In 1997 DESIRE produced a comprehensive review of metadata formats.[10] In the UK, UKOLN (The UK Office of Library Networking) has a Metadata Group that conducts ongoing research into metadata formats, especially in relation to library cataloguing formats such as MARC. Their Web site offers software tools for handling metadata and information on mapping between metadata formats.[11]

Strategy and Sustainability

It should be noted that the gateways described in this article are all large-scale projects with significant funding behind them. Many small-scale and/or volunteer-effort gateways have emerged on the Internet, but with the explosion of the number of sites these have not been sus-

tainable and have either been discontinued or have only been able to maintain very limited coverage with many broken links and outdated descriptions. Gateways require considerable staff time and investment to develop and maintain. Only organizations that develop a long-term strategy for their gateways can provide a service that is sustainable.

European Gateway Initiatives

A number of Internet subject gateways have been developed in Europe and a significant community of libraries involved in gateways is developing.

United Kingdom - The Resource Discovery Network

In the UK a number of subject gateways are being funded by the UK government's Higher Education Funding Council and are organized under the Resource Discovery Network (RDN).[12] All the UK gateways are based in universities and involve input from librarians and information professionals:
– BIOME - Health and Life Sciences
– EMC - Engineering, Mathematics and Computing
– Humbul - Humanities
– PSIgate - Physical Sciences
– SOSIG - Social Sciences, Business and Law

The Netherlands - DutchESS

The National Library of the Netherlands (Koninklijke Bibliotheek) has built a subject gateway in cooperation with seven university libraries called DutchESS (Dutch Electronic Subject Service)[13] - a national gateway, covering all subjects.

Finland - The Finnish Virtual Library Project

In Finland the government's Ministry of Education has funded the large-scale development of national subject gateways. The Finnish Virtual Library project[14] was launched in 1995 and involves collaborative

work between eight university libraries.

Sweden - EELS

EELS [15] covers the broad subject area of Electronic Engineering. It is a cooperative project of the six Swedish University of Technology Libraries.

Denmark (and other Nordic Countries) - NOVAGate

NOVAGate covers forestry, veterinary, agricultural, food and environmental sciences and is produced by the libraries of the NOVA University in Denmark, Finland, Iceland, Norway and Sweden.

The DESIRE Project

DESIRE is an international project funded by the European Union. The project aims to facilitate use of the World Wide Web among Europe's research community and one of the ways it is doing this is by developing and promoting the Internet subject gateways model. SOSIG, DutchESS and EELS are all partners in the DESIRE project and have been working with other gateways (including the Finnish Virtual Libraries project and NOVAGate).

DESIRE Workshop for Europe's National Libraries

There is considerable scope for the library community to be involved in Internet subject gateways. As illustrated in the gateways described above, many libraries in many countries are already seeing work on gateways as an important part of their remit. Once a country has a gateway structure in place, librarians from across that country can work collaboratively to build the collection. The subject gateways model offers strategic and standardised methods for doing this. DESIRE aims to support the development of new gateways in Europe, especially large-scale national gateways. In September 1999 there was a DESIRE workshop for the national libraries of Europe [16] "Building National and Large-scale Internet Information Gateways: A Workshop

for the National Libraries of Europe". At the time of writing, 17 European national libraries have signed up for the DESIRE workshop, and together we hope to make some important steps towards building a European network of gateways.

As the Internet continues to expand so quickly it is clear that no single gateway or country can hope to catalogue all the Internet resources available. A distributed model is required, where each country takes responsibility for describing the high quality resources available on its national network. Imagine the scenario where librarians from every country work at building a gateway to the best of their national Internet resources. Imagine then, that it is possible to cross-search any combination of these gateways - to find high quality Internet resources from around the world. In fact, the technologies and standards already exist to make this vision a reality. What still requires a lot of work is the development of the human networks that can maximize the potential of these standards and technologies - and the library community is perfectly placed to take up this challenge! Building an international network of gateways takes time, but the library community has both the expertise and the commitment to develop these valuable Internet search tools.

Distributed Teams of Librarians

Subject gateways provide a successful model for involving the library community in Internet resource discovery. Existing gateways have invested effort in developing systems that support the work of distributed teams, so that librarians can work on a gateway from anywhere in the world as long as they have access to a networked PC and a Web browser. Distributed Internet cataloguing means that libraries can contribute to a shared service, rather than having to each build a local service. This is an efficient way of working - it avoids duplicated effort and collaboration means large-scale gateways with much

better coverage can be developed. Many of the gateways described above benefit from the input of a distributed team of librarians. A DESIRE report "Distributed and Part-Automated Cataloguing" [17] describes the different models being used by existing gateways. The ROADS software supports distributed cataloguing by providing a Web interface to the database. Records can be added, deleted or edited remotely. All this work can be done via the Web - the teams can work from their own offices using their own workstations and fit this "Internet librarianship" in alongside their usual work in the library.

Distributed Databases

The technologies also exist to support cross-searching of distributed databases. Interoperability has been the focus of much research by DESIRE and ROADS, and other communities. If different databases of metadata records can be cross-searched this offers the potential for different communities to work at describing different sections of the Internet and for end-users to cross-search all these collections simultaneously. On a national level both the UK gateways and the Finnish Virtual Library project are working on cross-searching distributed gateway databases. The end-user remains blissfully unaware of the complex organization behind their search - from their point of view they are making a single search from a single Web page and get a single page of results.

SOSIG and Biz/ed have already implemented cross-searching into their working services. When users search SOSIG they are, in fact, also cross-searching the Biz/ed database - results from the two databases are returned on the same page. The technologies used to achieve this are described in a paper published in Dlib magazine. [18] Databases located in different countries can also be cross-searched simultaneously - DutchESS (in the Netherlands) has been working closely with SOSIG (in the UK) to set up a cross-search mechanism, so that

both the collections can be accessed simultaneously by users from both countries (and indeed elsewhere!). This is pioneering work and when it is in place, it is hoped the same mechanism will be used by other gateways to set up similar systems. Demonstrations of the cross-searching work being done by DESIRE and ROADS is available on the DESIRE Web site. [19]

Tools for Building Large-Scale Internet Subject Gateways

DESIRE is developing tools and methods for the development of large-scale Internet subject gateways. It is also working with both library and Internet standards organizations to develop standard practices for developing gateways, to ensure that they are interoperable and can work together to form large-scale, collaborative services.

The DESIRE Gateways Handbook. In October 1999 DESIRE published the *Information Gateways Handbook* - a guide for libraries interested in setting up large-scale subject gateways of their own. The Handbook is made freely available on the WWW and describes all the methods and tools required to set up a large-scale Internet subject gateway. It draws upon over three years of research into subject gateways and includes case studies and examples from many of the gateways described earlier in this article. It is hoped that the Handbook will assist other countries to set up their own national gateway initiatives so that more libraries and more librarians can begin to play a role in Internet resource discovery.

ROADS. ROADS[20] is an open-source set of software tools which enable the set up and maintenance of Web based subject gateways. It was developed as part of the UK's Electronic Libraries Programme but is now freely available for anyone to use. The software includes the database technology required to set up a gateway, the administration centre required to facilitate remote cataloguing via the WWW and everything else needed to run a gateway. Many of the gateways

described above use ROADS, notably SOSIG and the Finnish Virtual Library project. The ROADS open-source software toolkit is being produced by a consortium of developers with expertise in network-based resource identification, indexing and cataloguing. This has resulted in a standards-based approach to software development, making it compatible with current and developing indexing and cataloguing requirements. In addition, there is ample documentation and online support for people interested in using the software for either experimental purposes or service provision.

IMesh: The International Gateway Community

IMesh is a collaborative network, involving key players in the world's subject gateway community (not only those in Europe). It is likely that IMesh will be the key player in future gateway developments internationally.

IMesh was formed as a result of a meeting at the Second European Conference on Research and Advanced Technology for Digital Libraries, held in Crete in September 1998, attended by 25 delegates from 15 countries. One of the main aims of IMesh is to explore the potential for collaborative development of gateways internationally. It would require significant investment of effort and resources for a single country to attempt to create a gateway that pointed to the best of the Internet from all countries, in all languages in all subject areas. The IMesh group is looking at ways in which the effort can be shared through international collaborative agreements. Many of the technologies required for cross-searching different gateways and for remote cataloguing into gateways already exist. What is lacking is the strategic organization between gateways and IMesh aims to address this. In June 1999 the first IMesh workshop was held in Warwick in the UK and was attended by gateway providers from around the world. A report of the meeting is available on the

iMesh Web site.[21] An IMesh discussion list exists and those interested in international collaboration amongst subject gateways are invited to join. The list provides an open forum for exchanging ideas and technology for promoting the subject gateway movement.

Future and Conclusions

In many ways the Internet is still a bit of a building site! Many things are still under construction, including the basic architecture of the Web. The World Wide Web Consortium[22] is still working on building a structure that can support resource discovery on the Internet. They have recently released the Resource Description Framework (RDF) model and syntax specification, which aims to provide a basic infrastructure on the Web to support the transfer and processing of metadata. This marks a new age on the Web as in effect, it allows anyone to "catalogue" a Web resource in a machine understandable way. Different people will want to use RDF in different ways - it is simply the structure within which different people can work. Gateways are working with the W3C to see how RDF can support these high quality metadata collections. Potentially, librarians could forge the same role for themselves on the Internet that they have had traditionally - as third party information providers that end-users can learn to trust and rely on when searching for information.

Although the structure for the Internet library is not yet complete, it does not mean that librarians have to wait to start building their Internet collections. The human networks required to effectively catalogue the Internet will take many years to build. Libraries can work on creating metadata records for Internet resources and on finding their place in the metadata community. They can also start becoming familiar with the new metadata and Internet cataloguing standards so all the records are compatible. Although this paper was presented to the open meeting of the IFLA

Section on Information Technology, in many ways the technologies are the least of our worries - it is the human factor that now requires significant development.

In this article I have described services that already involve input from large numbers of libraries and librarians. Perhaps the IFLA community can help us to take this work forward and to promote Internet Librarianship as an important new role for the profession.

References

1 SOSIG (The Social Science Information Gateway); http://www.sosig.ac.uk/

2 AltaVista; http://www.altavista.com/

3 Yahoo; http://www.yahoo.com/

4 DESIRE; http://www.desire.org/

5 IMesh; http://www.desire.org/html/subjectgateways/community/imesh/

6 The Open Directory Project; http://dmoz.org/Computers/Internet/WWW/Directories/Open_Directory_Project/

7 Selection Criteria: Examples; http://www.desire.org/results/discovery/cat/selectex_des.htm

8 Internet Detective; http://www.sosig.ac.uk/desire/internet?detective.html

9 The Role of Classification Schemes in Internet Resource Description and Discovery; http://www.desire.org/results/discovery/cat/class_des.htm

10 A Review of Metadata: a Survey of Current Resource Description Formats; http://www.desire.org/results/discovery/cat/meta_des.htm

11 UKOLN Metadata Web Page; http://www.ukoln.ac.uk/metadata/

12 RDN, the Resource Discovery Network; http://www.rdn.ac.uk/

13 DutchESS (Dutch Electronic Subject Service); http://www.konbib.nl/dutchess/

14 The Finnish Virtual Library Project; http://www.uku.fi/kirjasto/virtuaalikirjasto/

15 EELS; http://www.ub2.lu.se/eel/eel-home.html

16 Building National and Large-scale Internet Information Gateways: A Workshop for the National Libraries of Europe (DESIRE);http://www.desire.org/html/subjectgateways/workshops/workshop1.html

17 Distributed and Part-Automated Cataloguing (DESIRE); http://www.sosig.ac.uk/desire/cat/cataloguing.html

18 Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach (D-Lib Magazine, January 1998); http://www.dlib.org/dlib/january98/01kirriemuir.html

19 DESIRE Cross-searching Demonstrator; http://www.desire.org/ROADS/cgi-bin/search.pl

20 ROADS; http://www.ilrt.bris.ac.uk/roads/

21 IMesh: International Collaboration on Internet Subject Gateways;http://www.desire.org/html/subjectgateways/community/imesh/

22 W3C (World Wide Web Consortium); http://www.w3.org/

- Communication and information literacy – (individual level)

The activities for the digital library in the first bullit are of course concerned with e-publishing and e-learning. The second bullet covers the library's role in enabling students to function as communicating participants in the knowledge society. This requires both information literacy (and i-skills) and the ability to express oneself and communicate using some of the many web 2.0 communication facilities.

Developing a new role for the library

The shift is service concept from "what you see is what you get" to "what you need is what you get" could help alleviate the risk of library bypass.

This requires the library to devote more resources to user studies, user involvement including developing a range of services utilising web 2.0 functionality.

Web 2.0 has two major advantages from a library point of view. As it transforms communication from being bilateral (user-library) to being multilateral (user-user-library) it saves resources for instance in relation to reference services as users can answer each others questions and share advice on good information resources. The other very important advantage is that it gives libraries' a unique opportunity to analyze and understand user's information searching behaviour making it easier for libraries to continuously develop relevant library services.

Information provision

The very traditional role of information provision has undergone very significant changes. The libraries have recognised that many of the processes involved although central to the library role does not add particular value to the end result. In line with traditional commercial wisdom for such processes the libraries have outsourced, consolidated and developed partnerships.

Many libraries have outsourced selection, acquisition and processing of materials to vendors. In Denmark the cataloguing of materials are shared on a national level or outsourced to vendors. DEFF handles negotiation and administration of electronic subscriptions and licenses. Much of the it-administration or development is outsourced to vendors or other libraries.

The processes surrounding lending are to a large extent outsourced via self-service to users. This includes ordering and reserving books, checking out and returning books and in some cases even being directly involved in the acquisition of books. We are fast approaching a situation were the users can check out and return the books in a library without staff involvement. The information provision in the form of printed books is one area where the library's traditional role of information provision has indeed become a commodity service.

Digital library services

There is a tendency for digital libraries to focus on the task of information provision as not merely an important task but the task that justifies the library's existence. This often means that attempts to consolidate or standardize the task are viewed as threatening to the individual institution. It is considered very important for the individual institution to be visible to the end user both in the role of information provision and via the user interface. This sometimes makes it difficult to collaborate on developing new services in this area.

Fortunately Google and Amazon have introduced a new standard for digital information provision that has increased libraries motivation to co-operate in more areas. Two of the most important drivers for cooperation and consolidation in information technology are cost and complexity. Libraries have now realised that they have neither the skills nor the resources to deliver the functionality that users have come to expect from Amazon and Google - unless they cooperate.

Libraries have also come to recognize that users expect to be able to access all available information from one search box and to get exactly the result they need because of excellent search technology. Most libraries' information provision has traditionally been centred on the library catalogue and this has been the starting point for the ambition to deliver a Google-like search interface. For many libraries this ambition demanded that all the digital information resources was catalogued and made available through the library catalogue. This strategy proved insufficient because the library system is ill suited to handle the task, the amount of digital information is too huge and because of the manual labour required.

An alternative approach is using federated search. This allows users to search multiple information services form a single search box but it normally utilises the search facilities of the individual services which means that it is difficult to provide an intelligent raking and presentation of the combined result.

The current approach under consideration in Denmark is to use integrated search. This demands that data in full text or metadata be aggregated and indexed in order to provide the same intelligent search and presentation to the users as Google delivers.

This approach has the advantage of presenting the user with very relevant search results but many of the information resources will not necessarily be immediately available to the user because the library's own information resources only constitutes a very small part of the information presented. The presentation of so many information resources external to the library makes automated document delivery services and complex authorisation, authentication and accounting (AAA) natural components of integrated search.

In the Danish context libraries are discussing the establishment of a mutual national data repository to be used as basis for the individual libraries' indexing and searching solutions. Other important areas of cooperation include AAA, document delivery, data on user behaviour and webservices to be used in adding relevant information to the presentation of the search result.

Concrete activities and service development

The different aspects of access control embodied in AAA are becoming increasingly important because the transition from print to digital means that "fair use" is replaced by contractual regulation of the use of digital material. The ability to control and limit the use of information resources is often a precondition of making the material available in the first place. There is a need for a standardized and unified system across libraries and vendors which preserve the privacy of the end user while at the same time enabling single sign on to many different information resources. In the Danish context the project working with Shibboleth are seen as a step towards such a solution.

Comprehensive and user friendly solutions for integrated search, automated and flexible systems for document delivery and a common AAA-system are important components of an efficient information provision. The functionality required for collaborative filtering, recommendation services, customization and customerization is fast becoming a minimum requirement for the user interfaces to information resources and DEFF is looking to develop this functionality as services in the common architecture.

These activities are supplemented by cooperation with Google Scholar with the objective of making the Danish union catalogue and other relevant Danish digital information resources available through Google scholar thus increasing the visibility of these resources to end user. A separate focus area in a Danish context is the digitization of Danish scientific and educational material. Legal difficulties and reluctance from publishers has so far blocked attempts to digitize these materials on a larger scale but initiatives like Google Print and the EU's i-2010 initiative have contributed to a gradual opening in this area.

E-publishing

DEFF's activities in the area of E-publishing started with projects on institutional repositories and the migration of journals to open access. Inspired by the e-framework proposed by The JISC there are discussions about whether the systems established for e-publishing are merging with e-learning. There is an increased focus on library support for the general knowledge management for the institution in all the

processes between import and export of information. These deliberations reflect the more general idea that the libraries take on a much more active role in supporting the information flow and knowledge management for institutions and individual users.

An important driver in this development is the need for systems not only to handle publications and preprints but the underlying datasets and part of the research communication as developments within e-science and e-research seem to suggest. Some libraries are expanding their activities into new areas such as developing research statistics and offering citation analysis and benchmarking of research activity. In addition to these demands in relation to research there is a demand for repositories for student papers and assignments.

A very important part of the work in this area is the technical and legal framework for the activities. The Danish political awareness of open access has been very modest but new initiatives are emerging and there is a Danish franchise of the creative commons movement. At the institutional level we are witnessing the formulation of guidelines for publications demanding that researchers deposit copies of preprints in institutional repositories.

E-learning

The increasingly important area of e-learning has presented libraries with a range of new challenges. The correct term for the activities is in fact blended learning. The term indicates that learning should take place both in the electronic and the physical environment. Blended learning is however also a very appropriate term to describe a development very students are increasingly learning from each other outside the class room and very library and teaching staff are cooperating to facilitate this learning. This cooperation has proved valuable but also difficult and raised discussions as to whether the role of libraries and librarians should be limited to information provision. It is our view that the librarians should develop the relevant new competencies to cooperate with teaching staff in facilitating learning. As indicated in the model above this is one area where the library could add substantial value to the activities of the parent institution.

The projects supported by DEFF in this area have to a large extent sought to develop the information provision in regard to e-learning. This includes integration of library services with the institution*s virtual learning environments and the provision of information resources for particular courses. Another important area has been new way to develop the student's information literacy.

The work in relation to e-learning involves a host of legal tasks such as licensing the use of electronic information resources for e-learning, negotiation agreements for digitisation on demand and formulating institutional policies regarding the ownership of learning materials produced by teaching staff.

Organisational aspects - Institutional partnerships

There is probably not one right way to develop the library at a particular institution. The local libraries develop according to local competencies and demands from the parent institution. The more conceptual discussions of the role of the library could nevertheless prove important because it helps alleviate fears of the imminent demise of the library and tries to point to new tasks for the library.

These discussions might also clear the way for more rational cooperation between libraries and a better division of labour between the local library and national digital library services. In Denmark there are discussions as to whether libraries could divide their services into back office and front office functions thus allowing the further development of shared services and outsourcing. These discussions are particularly important in the area of information provision where the concept of a local digital library is being challenged by increasing technical complexity and user demands.

The concept of the library role as being one of facilitating knowledge creation is sufficiently broad to allow individual libraries to develop services in new areas where they can probably add more value than trying to develop the traditional role of information provision on the own. The new tasks would allow the libraries to develop a new and closer relationship to teachers, students and researchers. Such a role would imply a closer integration with the parent institution where the library would be perceived less as a monolithic institution and more as a network of services. In order to fulfil this role it would be important for libraries to cooperate nationally or internationally in the area of information provision and to collaborate in developing the new services and competencies in new areas. As an organisation of research libraries, DEFF sees these discussions as essential to the continued development of the individual library and as a very important contribution to deciding which digital library services to develop on a national level.

Staffing the digital library

Naicheng Chang
Associate Professor, General Education Center
Tatung University, Taipei, Taiwan
naicheng.chang@uclmail.net

Alan Hopkinson
Head, Library Systems and Bibliographic Services
Middlesex University, London, UK
a.hopkinson@mdx.ac.uk

Abstract

In the real world of academic libraries, the digital library world has had a profound impact on staffing. Academic libraries are facing huge pressure on their staffing levels at a time when digital libraries are being introduced.

- Who enables access to digital materials?
- Who creates digital materials?
- Who selects digital materials?
- Who promotes their use?
- Who trains users?

Digital libraries cannot be divorced from the ordinary library service. What skills do traditional librarians need to acquire? There is some opposition in various levels of the profession to librarians becoming more involved in some of these aspects. When computers were introduced into libraries there was opposition to librarians often from senior librarians to their being involved more. In some countries, librarians and computer professionals are separate disciplines and no one would have a foot in both camps. Sometimes this resulted in other groups such as computing professions stepping in.

What is the situation with digital libraries? There is little in the literature on training for staffing for digital libraries. Consequently, evidence from a recent PhD gleaned from research interviews of these digital libraries case studies will be included in this paper. This PhD uncovered a variety of different management and organizational issues and the large cost of personnel in the implementation and maintenance of digital libraries

Introduction

As part of a study on the impact of the XML (Extensible Markup Language) (Bray *et al.* 1998) in digital library development, studies were undertaken on three digital library initiatives, which were chosen because of their size and because they were examples in the global digital library community. These studies included research interviews conducted during visits in September 2002 to the three projects which were in different types of library (Chang 2006).

The PDL (Perseus Digital Library) is an instance of a research and development testbed (PDL n.d.). It represents the mission of one Humanities scholar to use technology to facilitate research and teaching in the humanities.

The University of Michigan DLS (Digital Library Services) is representative of an academic library. In June 2003 the DLS was renamed Library Information Technology (University of Michigan n.d.). Its mission is to support a virtual learning environment and preserve campus-wide materials for long-term access.

The LC (Library of Congress) was taken as an example of a national library. Its mission is to make its resources available and useful to Congress and the American people and to sustain and preserve a universal collection of knowledge and creativity for future generations. As part of its mission, it has established the NDLP (National Digital Library Program) (Library of Congress 1998).

This paper is taken from the results of the research interviews and their analysis.

The development of electronic information brings new management challenges to libraries. New tasks have emerged calling for different skills and new input of resources. These involve the revision of information strategies concerning planning, introducing, monitoring, and improving electronic services. Also, there is . growing demand for staff to be associated with the multiple skills related to the creation, acquisition, recording and management of data available in new or expanded areas of activities. In LC and University of Michigan, the digital libraries were part of a traditional library and staff could move from one role to another, from the traditional area to the digital or even a hybrid role. Perseus is different because it is the mission of one man who is working outside the traditional library sphere.

Digitization is much more than scanning data and digital libraries need more than just to advertise their existence. Markup has become the key to the efficient storage, location, promotion, retrieval, analysis and evaluation of information. People with knowledge of markup and its implementation, either computing personnel or skilled librarians, need to be recruited, or existing staff will need to have more training, particularly in XML.

The investigation of these management issues was undertaken by asking the following questions in the interviews.

- How has staff structure changed due to digital library development, especially
- XML-based? What are the implications for staff recruitment, retention and development?
- What kind of people do you need to hire to do the work, particularly tasks related to XML technologies?
- Do you provide staff training? What skills do they need?

Staff infrastructure for the digital library

Technological advances have brought great changes in library and information services. The changes are set to continue as libraries are expected to become fast moving, reorganizing and innovative (Arms 2005). It was discovered from the interviews reported below that in order to be able to incorporate well into the roles, . librarians require a wide range of new and enhanced hard skills, such as technical knowledge, and soft skills, such as vision for the future.

Perseus

The view of Perseus was that using digital resources to do research was beginning to have a great impact in the humanities. Basically, Perseus looked for people in the humanities with little computing experience and trained them to do research and to create electronic resources. This had one major impact on the way that Perseus structured the work; that is, wherever possible Perseus did its work in-house, preferring to use its own resources to train and support young scholars in the humanities rather than supporting outside professional contractors.

Michigan

Staffing in Michigan has grown with its accomplishments. Initial staffing was set at the levels that are necessary to provide a baseline of commitment to all areas, with growth expected for new formats and for extending DLPS (Digital Library Production Service) commitment to issues such as cross-collection and cross-format integration. At the time of the research visit, DLPS had approximately 24 full-time equivalent staff; a small part of them were short-term-funded people related to a specific grant. Michigan provided limited staff training to all staff but focused on digitization and encoding skills to staff that are in charge.

Staff in Michigan are grouped into two areas of work. The digitization group focuses on methods and formats and is responsible for production-level creation and conversion of digital library resources supported with markup technology; the information retrieval and architecture group build the digital library infrastructure to ensure a smooth delivery system. This group also works closely with the digitization group to

ensure the extremely high volume, high quality digitization operation. Many programmers within this group have areas of specialization including XML and SGML. This ensures a high degree of technical and format understanding in building online systems. The infrastructure work involves most areas of DLPS operations such as interface specialist, data loading and technical support for DLPS staff. It was impressive that DLPS had been working closely with the university librarians, exchanging information across units or within the Library as a whole. For example, library cataloguers are acting as DLPS's metadata specialists. Whenever necessary, DLPS would consult University librarians about naming collections or metadata mapping.

At the beginning, Michigan had more staff with a humanities background, but now they are recruiting more staff who have a computing background with interest in the humanities. Among them, some are librarians who have just got involved working in computing; some have a really complete programming background; and some a humanities degree. In the future, Michigan is planning to recruit more people with computing expertise and experience in digital library evaluation, especially relating to user needs, to cope with the difficult work of developing an increasingly complex digital library information infrastructure.

The Library of Congress

LC highlighted the need for changing the skills and emphases of the organizational changes in the library. Skills associated with particular data types such as XML and SGML have become increasing important to staff.

To support the NDLP, LC developed a well-organized staff infrastructure as follows (Campbell 1995).
- *Curatorial staff* Staff assigned to the curatorial divisions prepare and process materials to be digitized. Curatorial staff also performed on-site digitization of materials that include rare and fragile items such as early drafts of the declaration of independence and the Gettysburg address.
- *Core staff* NDLP core staff worked with the Library's divisions to prepare and describe the collections, verifying the status of copyright and seeking permission for use of the materials when appropriate, and digitizing the materials and verifying that they adhere to the Library of Congress's standards of quality. Digital conversion specialists in the central office provided project coordination and technical oversight. The more experienced specialists oversaw collection development and production, serving as team leaders and as brokers among the division and automation staff and contractors.
- *Infrastructure staff* Infrastructure staff were primarily information systems experts who built and maintained the automated systems that stored and provided access to the digital collections.
- *Educational services staff* The educational services staff focused on educational outreach for the use of the historical collections by school children between the age of 5 and 18. They researched user needs, talked to the education communities, evaluated technologies for delivery of digitized materials, coordinated collection selection and developed and supervised contracts.

During the interviews it transpired that a great deal of effort had been put into formulating new job descriptions as NDLP was a vast and new task. LC found that personnel who could bridge the gap between traditional librarianship and technical skills such as those who could build delivery systems were highly valuable and in great demand.

At its peak, NDPL employed as many as 100 people with various professional skills related to the needs of building a large digital library. LC provides several kinds of staff training packages both on-site and off-site and has sent many staff to training courses.

It is worth noting that LC staff have been actively involved in activities related to XML technology. For example, when the author was interviewing in LC in September 2002, there was a workshop being held there by a commercial institution on the subject of XML-based TopicMap technology, which can be used in information retrieval in a digital library. LC core staff were interested to know more about XML because they knew that they needed to prepare and know the technology well to decide how XML could be helpful in the digital library development when the technology is ready for them.

76

Challenges for digital library staffing

Technical challenges in the digital library have, in general, already been recognized, while the non-technical challenges encountered by digital library developers have been proving to be more elusive, complex and profound. The digital library research community is increasingly concerned with the need to base the design of digital libraries on the work of the community it supports (Star and Bishop 1996). The social impact and influence of digital libraries have raised active discussion on many occasions over recent years, such as at conferences on digital libraries (ACM n.d.). The challenges fall into several aspects across the digital libraries spectrum, including organizational co-ordination between a number of partners, licensing, and copyright concerns regarding online materials, user studies and so forth.

Pinfield (2001) examined the new roles and responsibilities that librarians are now involved in as new era digital librarians. He thought librarians would act as multi-media users, intermediaries, enablers, communicators, project managers, trainers/educators, evaluators, metadata producers; team players negotiators, innovators and fund-raisers.

Our view is that in Pinfield's list, innovation would be the key requirement for librarians in the fast-changing environment of the profession. Institutions are not looking for people doing routine work but for those who have knowledge of future trends in their professions. Therefore, our view is that librarians in general, following the lead of those in the Library of Congress, are wise to see the future in XML. The results of our investigations of real world library posts discussed in the following paragraphs support our vision of this.

Research on the perception of the importance of XML by managers and trainers

A division of the American Library Association, the ALCTS (Association for Library Collections and Technical Services) Continuing Education Task Force undertook a survey as a response to Action Item 5.3 of the Library of Congress' action plan "Bibliographic Control of Web Resources". This survey discovered that 73% of respondents thought that cataloguers needed knowledge of XML today and would for the near future at least (ALCTS Continuing Education Task Force [Action Item 5.3] 2003).

In order to examine the theory of the need for XML knowledge as against the perceived need for XML as reflected in library job advertisements, at the beginning of September 2004 (since September is one of the peak months in the year for advertising jobs), we did an investigation using Web job listings in the UK, the US, and Taiwan. We targeted academic librarian jobs and computing jobs in the academic sector. Furthermore, we investigated whether in those three countries XML was part of the curriculum in library schools and if there were reports or activities related to XML from the library associations in the three countries. Below are our findings and analysis.

In the UK, between July and the beginning of September of that year, none out of the 31 advertisements for professional or managerial jobs listed in jobs.ac.uk (2004) needed XML during that period; 2 out of 74 put XML required in the job description for professional or managerial computing jobs; 3 out of 14 need XML for computing technician jobs.

In the US, there were three out of 151 librarian jobs listed in the Association of College and Research Libraries across the country that stated that they needed XML in the time period 25 May to 1 September. 15 out of 153 academic computing and information jobs listed in EDUCAUSE (a US association which promotes the intelligent use of information technology in higher education) needed XML in the time period 2 July to 7 September 2004.

In Taiwan, between 9 January and 3 September, none of the 25 librarian jobs listed in the LAC (Library Association of China) job listing needed XML, but LAC gave 3 training courses in XML in 2004 under Digital Archive and E-learning schemes; none out of 7 computing jobs listed in the National Youth Commission Website (a government job website) needed XML in the academic sector during the time period 17 December 2003 to 3 September 2004.

77

In the library sector, most of the jobs which required XML are in areas such as digital library projects, electronic services and metadata while in the academic computing sector, knowledge of XML was stated as being needed more for computing technicians than for the managerial level. However, it is interesting to note that for managerial level library jobs, institutions were looking for candidates who are innovative, creative and able to provide strategic direction and vision for the libraries; they must be able to provide both traditional and innovative library resources and services, possess an informed vision of the library in the 21st century, knowledge of new trends and emerging technologies in information services, and be capable of identifying future needs and directions in an electronic environment. For instance, for the post of bibliographic systems manager, candidates must have substantial innovation and creativity to identify new ways to use bibliographic services to delivery library resources to users' desktops; for a job as head of cataloguing and acquisitions, they must have knowledge of current cataloguing standards for all formats and awareness of emerging trends and technologies in technical processing; for professional library jobs, institutions are looking for a knowledge of IT issues, ability to learn and apply new technologies to improve work operations, ability to handle multiple responsibilities in a changing environment, non-traditional thinking with regard to library collections and services, ability to meet professional standards and competencies and so forth. For Digital Projects Librarian, they must be familiar with structured markup and knowledge of metadata standards and best practices; for reference librarian, they must be aware of current trends and emerging technologies in the delivery of reference services; for university librarian, they must possess a wide understanding of the current technological information environment and an innovative vision for the library as a partner in developing approaches towards access to information; for an information systems librarian, they must have knowledge of metadata standards and best practices. These words do not explicitly state what experience would qualify a candidate but, in a sense, they could be XML-related technologies. Incidentally, statements of job requirements in Taiwan are customarily relatively short.

We also noticed that in the library job sector those jobs relating to automation or digital libraries did not mention XML though some mentioned knowledge of markup language and best practices in digital projects. The lack of mention of XML may be due to a lack of understanding of it by library managers, or the fact that knowledge of best practices as stated in the job descriptions covers more than one technology and XML is intended to be one of them. We note that there is a relatively small library job market in Taiwan. Librarianship in Taiwan awards a bachelor's degree to students who have spent four years studying librarianship and information technologies. Students spend two years for a master's degree for the study of advanced librarianship subjects. Librarianship students find jobs in various fields as the line is blurred between librarianship and information technology.

Library schools have not recognized the need for XML skills. According to their course descriptions, in the UK, only one (University College London) out of 8 library schools provided an XML course; in the US, two (University at Albany, SUNY and University of California, Los Angeles) out of 50 library schools provided an XML course; in Taiwan, one (Hsurn Chuang University) out of nine library schools provides an XML course. On the other hand, the concept of XML has been introduced in several courses, for example in electronic publishing (Sheffield University and University of Illinois at Urbana-Champaign), in document engineering (University of California, Berkeley and University of Illinois at Urbana-Champaign), in technologies in web content management (Syracuse University), in information organization and access (University of Illinois at Urbana-Champaign), and in information technology tools and applications (San Jose State University), in access systems for archival materials (University of Michigan) and in taxonomy, classification, and metadata (University of Washington); in library school workshops (Kent State University and Tamkang University in Taiwan); or has been experimented with as faculty research projects (The University of Strathclyde, University of Illinois at Urbana-Champaign, Tamkang University and many others).

Overall, XML appears in more course descriptions at the University of Illinois than anywhere else. This is probably because teaching staff there have more knowledge of XML. We noticed that in Taiwan there are at least two doctoral-level computing specialists in each library school, giving courses in database management and programming related courses; therefore, there would be no problem for the departments to provide XML-related courses.

In the UK, the CILIP (Chartered Institute of Library and Information Professionals) has been announcing more training workshops on library and information technology in which XML-related initiatives such as markup language, schema, RDF (Resource Description Framework) are among the topics (CILIP 2006). CILIP also published Hey's (2004) article on the vital role of academic librarians as metadata experts and digital curators in the digital age; he mentioned that librarians should be well versed in current new technologies such as XML. He felt it is important to mention in this context where the readership consists of librarians at all levels that senior librarians represented by the Consortium of University Research Libraries and the Research Support Libraries Group had shown little vision in this direction though he conceded that JISC was putting effort into these training tasks.

In the US, since 2000 the Library and Information Technology Association at ALA has selected XML related initiatives such as MARC XML as annual top technology trends (ALA 2005). Library Association of the Republic of China in Taiwan provides irregular XML training courses. This could be because there is a five-year nation-wide national digital archive programme and the training sessions are designed to support that programme.

To conclude our investigation, although XML has not been recognized as a core skill in library jobs or as part of the core programme in library schools, nevertheless, library associations have identified XML as an important technology trend that needs to be monitored carefully. Indeed, as Felstead (2004) pointed out, XML was already being extensively used in a substantial number of library systems, though the use of XML was transparent to the librarians using the systems. We suggest that it would be advantageous for librarians to have knowledge of XML even if they do not work directly with XML; and library schools could provide selective courses on XML.

Career development for the digital library

Very often, digital libraries are part of larger organizations; therefore, there will be a need to review the fundamental organizational structure of how staff are obtained, trained and retained in order to carry out the wide range of library services in the digital age. Staff retention is important and it is necessary to have a good career structure in place, which is going to be difficult in most environments where academic digital libraries are being developed. Young staff will want to see a career path ahead of them with the opportunity to earn more money as time goes on. A digital library where the only benefit for the workers is pride in their work will not help in their retention. Being in a pioneering field with skills in a tool like XML which is used in the commercial world will mean that young workers will be in a good position to seek posts in the commercial world rather than in the academic one. Small organizations will find it difficult to provide a career path, and large organizations like the Library of Congress will also have the problem that employees working on digital libraries may find career promotion within the organization, but outside digital libraries, to the detriment of digital library development.

Inserting markup in a text is an act of interpretation (Hockey 2000); this will need domain specific specialists to deal with the creation of markup in the documents. As Perseus is a humanities-based digital library, there is a rising demand for corpus editors who combine technical and traditional humanistic expertise to accomplish particular domain-specific goals. The main job of corpus editors is expected to be to establish a reasonable level of precision that balances scholarly standards, and also to document clearly the level of precision employed to meet users' expectations when reading documents in a corpus (Crane and Rydberg-Cox 2000). Crane remarked when interviewed that Perseus had been fortunate in being able to recruit true corpus editors who successfully created the Greek and Latin texts that tied together scalable methods of tagging and a specialized knowledge of classical languages and literature.

Perseus has a very different staff training model from the other two case study libraries. Perseus exposes young scholars in the humanities to the new technology in a dynamic learning environment. Core staff (about three to five people at a time) are expected to learn XML as a basis. Some staff focus on the most demanding and rigorous software engineering; others may concentrate on the humanistic content more directly. By and large, most of the Perseus personnel are computing specialists to one degree or another, and everybody wears

many hats; that is, Perseus staff are able to manage the whole workflow production from content creation to programming and putting the content on the screen.

We noticed that the Perseus' managerial staff infrastructure is heavily constrained by the budget. Compared to the other two case study libraries, the Perseus staffing group is small but dynamic and efficient in terms of the size of the collection they have created. Our view is that this can be regarded as one of the outstanding features found in the Perseus environment. Staff in Perseus do computing and creation and they thought it was an advantage for them because there was no communication gap, and hence greater efficiency. On the other hand, Perseus staff are not librarians and do not have support from the University librarians when dealing with the library area of work. For example, cataloguing librarians could be the best metadata people to be consulted. We suggest that this could be a disadvantage for Perseus. If Perseus was not just a departmental activity but was formally integrated into the university library, this problem could be resolved. We have found that it is quite common for digital library projects in institutions to exist completely apart from the traditional library to the detriment of both parties, though they do not admit it explicitly. For example, in the Website of the Library of Waikato University in New Zealand (University of Waikato 2006), there is no hint that the New Zealand digital library is produced in the same university's department of computer science.

Both Michigan and LC have strong support from their libraries with overlap in staff. Perseus did not have the problems of organizational change as it was created as a new service but it has no contact with the university library. From the three cases we studied, we noticed that there is a substantial difference between institutions in staff development schemes, as they are not equally positioned to supply training opportunities. Mostly, this is subject to budgetary restriction and lack of staff time. Additionally, the reorganization of the library staff into a team-based learning organization would require libraries to think about how, by whom, and in what combination certain tasks can be done in order to fulfil the tasks successfully. LC and Michigan are examples. Furthermore, we also noticed that librarians in Michigan and LC with many years of experience act as core staff, supervising staff specialized in different areas and controlling digital library processing and future development. The core staff are attending more XML courses, evaluating the possibility of implementing XML.

To develop an ongoing and innovative staffing approach well, there has been a positive encouragement to publish, offer conference papers or become engaged in national and international activities. Perseus was a good instance of this case; our view is that this could be regarded as one of the successes of the Perseus marketing strategy.

It is interesting to know that our case studies needed more computing staff as time went on. In the case of adopting RDF technology in digital libraries, we see that Perseus staff are computing people and collection creators; therefore, they could decide to use RDF on their own initiative. On the other hand, LC core staff coordinated and made decisions on things but they were not computing people; although they learnt from meetings about the advantages of RDF, they were not interested. Michigan librarians were not familiar with RDF. This indicates that librarians are in charge of the digital library development but as they lack computing knowledge, they may not be able to take advantage of the cutting edge technologies in the first place. This again highlights the importance of the librarians' skills in the digital age as indicated by Pinfield.

XML is a new technology for the Web and it could play its part in every library operation. Librarians in the future will play a mediating role between the computing professions and the users. Knowledge of current standards and newly emerging technology trends such as XML will be a beneficial skill for librarians while looking for jobs either in the library sector or information-related sectors. Library schools could therefore contribute to the acquisition of this professional knowledge by covering these subject materials in their curriculum. As Hey (2004) suggested, librarians should be aware of the technological trends and be prepared, in order to compete and survive in the ever-changing environment.

Are these findings universal?

The studies were all of libraries in the US since when we began our research there were few developments elsewhere. We have detected from personal experience in the UK and Asia that in some areas there is opposition in various levels of the profession to librarians becoming more involved in the management and

technical activities associated with digital libraries. When computers were introduced into libraries there was opposition to librarians often from senior librarians to their being involved in the technical aspects. Senior librarians may feel deskilled in face of a digital library whether it be in developing it or being a user of it or assisting library users to use it. There are instances where senior librarians do not want their younger colleagues who are much more computer literate to be involved in these activities, preferring to hand them over to computer professionals who in practice may have little understanding of the principles of information retrieval. In some countries, librarians and computer professionals are separate disciplines and no one would have a foot in both camps. So, there are many reasons why this may have resulted in other groups such as computing professions stepping in. Clearly, there is ample material for a further study.

Conclusion

The results of the research indicate that staff development is a significant issue in a changing library environment as well-equipped professional librarians are key resources to developing and maintaining a high-quality digital library. To ensure the maximum exploitation of staff skills, efforts need to be made to determine ways of assessing skill level requirements and performance and training effectiveness. A continuation of specialized and dedicated training periods will be carefully planned to meet the needs of staff re-skilling. All of these developments will have their ultimate aims, that is to enhance library staff's ability to deliver a more responsive and effective service to the direct benefit of library users. It would be helpful if library staff had XML knowledge to cope well with library work as they know that XML is already in their work or is anticipated to be involved in future. On the other hand, knowledge of XML could lead to a librarian being able to seek a much better paid job in the commercial sector, since XML has many commercial applications. This could apply equally to computing staff.

Digital libraries are applying an increasingly sophisticated use of information technology in their digital library systems. Such libraries need more computing professionals to manage the systems, especially people with XML knowledge. We have seen through our investigations that it would be helpful if librarians had some knowledge of XML, as XML will be increasingly involved in the whole process of library operations. Librarians in the digital age will need to blur the distinctions between library and computing professions and, as far as librarians are concerned, change their role from that of traditional book-keepers to those of "cybrarian", working closely with computing personnel. This could also apply to the scenario of Perseus where Perseus management raised the issue of the dearth of institutional structures in place to produce corpus editors, who know enough about both computational algorithms and the documents in the corpus to adapt these techniques for the corpus in question. Digital libraries are looking for staff who are forward-looking (that is, visionaries), and can particularly act as team partners to explore and advise institutions on key issues in developing digital libraries. More attention must therefore be paid to the retention of staff with XML knowledge, which is valued in the commercial world as well as in academic digital library development.

Bibliography

ACM. n.d.
ACM International Conference on Digital Libraries
http://www.acm.org/ (accessed on 6 January 2006)

ALA. 2005.
Top Technology Trends by Topic
http://www.ala.org/ala/lita/litaresources/toptechtrends/ toptechtrends.htm (accessed on 8 April 2006)

ALCTS Continuing Education Task Force (Action Item 5.3). 2003.
Cataloging for the 21st Century: a Proposal for Continuing Education for Cataloging Professionals

Using SOSIG to support Social Science Teaching and Research

Abstract

The Internet can provide an invaluable resource for supporting education and research in the social sciences; it offers access to people, data and resources on a hitherto unparalleled scale. However, even with the advent of the more user-friendly World Wide Web, it is as yet a far from ideal work environment. The sheer enormity of information available and the corresponding lack of organisation of this information can prove an effective barrier to potential users.

The Social Science Information Gateway (SOSIG) allows social science researchers and practitioners easily to discover and access relevant high-quality networked resources and services world-wide including data archives and statistical software. It also provides a comprehensive list of relevant UK-based sources. SOSIG points to hundreds of resources on subjects ranging from Anthropology to Statistics. All the resources that appear on the gateway have been catalogued and described to aid users.

Introduction

The basic definition of the Internet is a global network of networks, a physical infrastructure linking hundreds of thousands of computers and enabling, via various software tools and information 'protocols', the exchange of information between them and between the more than 30 million people across the world. With electronic mail, discussion and news groups it provides a method of communicating with people locally, nationally or internationally on an individual or group basis in a way that is both fast and economical. The Internet also allows individuals to use services and systems - such as library catalogues and data archives - 'hosted' on computers anywhere in the world when they do not have access to such services locally.

The culture of the Internet is still very much one of co-operation and a readiness to share knowledge and expertise with others, so whatever your interests or specialisms there are likely to be groups of people who share those interests on the Internet. Traditionally academics have shared resources such as working papers, articles, library catalogues, data archives and so on and the number of these has grown as publishing information on the networks has become easier. However, as more commerical institutions and businesses are becoming involved in the Internet a whole new range of information and services are also appearing.

The Internet - this global network of networks and the means of exchanging information - has been around for 20-30 years, but the advent of the World Wide Web (the set of tools and standards which allow images, sounds and animations to be transmitted and offer a 'graphical user interface' to information) has made publishing and accessing information much 'friendlier' and, in theory, much easier. When Internet access first escaped from the domain of the computer experts in this way many others in the academic community "gave it a try" and found it wanting. The most common complaints were the lack of availability of useful materials and the difficulty in locating the few that did exist. These problems persist with the continuing rapid expansion of the Internet: imagine a vast second-hand bookshop with a constantly expanding and changing stock but with little more than the spines of hundreds of thousands of books visible. *Rose Growing* and *Essays on Stress* jostle on the shelf next to *Developments in Cognitive Psychology*. Just like browsing the Internet, random selections from the shelves may uncover something up-to-date and worthwhile, but may just as easily disclose nothing more than outdated or frivolous comics on subjects of little or no interest. The time and effort taken to clamber to the top shelf to retrieve *Essays on Stress* will have been wasted if eventual perusal of the back cover or a quick flick through its pages reveal it as a collection on stress fractures in bridges (or worse still, simply a list of titles of other such works) rather than on the psychological and physical effects of occupational stress you were looking for. A recognition of these difficulties led to the eventual establishment of the Social Science Information Gateway project in the UK.

82

Background to the Project

In 1992 the UK's Economic and Social Research Council (ESRC) appointed a Networked Information Support Officer to examine the potential for the use of networked information amongst the UK social science community and to encourage further development. There was a perception that the social science community were not reaping the benefit from networked resources in the same way as their colleagues in the natural and physical sciences. Through holding training sessions and workshops during these early days, it became clear that, whilst researchers could eventually locate resources where guided by an instructor and assisted by training documentation, the picture was very different when they tried to incorporate these newly-acquired skills in their day-to-day work. Disappointment and frustration soon followed the exhiliration and excitement of the first trips on the Information Superhighway. Neither the inclination nor the time were available for extensive browsing in an attempt to find something useful. It was decided to try to provide the social science research community with an easy way of locating electronic information and data which could be used in their work. The project which grew from this idea was the Social Science Information Gateway (SOSIG). A pilot service of the gateway became operational in June 1994.

Use of World Wide Web

The primary objective of the project was to establish a 'one-stop-shop' of UK social scientists, to connect the user seamlessly to relevant resources regardless of their location. World Wide Web was chosen as the software tool to deliver this solution for various reasons: it provides the consistent and easy-to-use interface that was required and was the fastest-growing NIR (Networked Information Retrieval) tool at the time, with development work being carried out in Europe and the United States. It also allows access to many other processes and protocols such as gopher servers, ftp sites, telnet sessions and many more of the earlier systems and services hosting useful social science data. The subsequent development and popularity of the Web for delivery of both academic (teaching and research) and commercial and recreational information serve to vindicate that choice now the Gateway is well-established and others are following the model.

Structure of SOSIG

There are over one thousand links to social science resources on SOSIG at present, with new resources being added on a regular basis. Users are presented with an uncluttered 'home' page with self-explanatory and simple-to-use buttons (with text alternatives) allowing browsing of resources by subject heading (subjects now range from Accountancy and Anthropology to Sociology and Statistics), Universal Decimal Classification (UDC) order or searching of the SOSIG database of resource information. Users have the option of connecting directly to resources they locate through SOSIG or reading descriptions and keywords **before** making the decision whether or not to connect. Unlike many other gateways or subject listings, the project tries to maintain a level of quality control. Resources do not appear on SOSIG until they have been filtered, recorded, classified and **described**.

Identification and Collection of Resources

A variety of sources are scanned for relevant resources including:

- Mailing lists and newsgroups
various lists and newsgroups are scanned for announcements of interesting social science resources, some of these groups are subject related but others are general lists set up specifically to monitor and disseminate information about new network resources.

- Printed guides and catalogues
these often provide a good starting point for identifying resources within a particular subject area.

83

- Networked search tools
  there are a number of tools or robots that aim to build indexed catalogues of resources available over the Internet, such as Veronica, Archie and Lycos and the ever-growing number of other Web search engines and virtual libraries.

- Other networked services
  following links from other services on the Internet.

The project also has a number of volunteer LISTeners. These are generally subject specialists in the social sciences who can advise on the quality of resources. In addition to recommendations for the LISTeners the project also regularly receives recommendations from the users of the service. An on-line form is available via the 'Add New Resource' button on SOSIG's home page for users to e-mail suggestions and additions (as well as occasional corrections) to the gateway. These are subject to the same quality checks before they are added to SOSIG. Regular, automated checks also avoid 'dead' links which occur when sites move leaving no forwarding address or vanish altogether - although as one of the selection criteria for SOSIG is 'stability' this happens less often than elsewhere.

Filtering Resources

An important role for SOSIG is to filter out 'junk' - resources that are of little or no use to our users. Resources are chosen according to selection criteria that include areas such as relevance, reliability, stability and currency. The popular Web search engines have much larger databases and often retrieve thousands of 'hits' which at first glance look exciting, but which are often in most cases irrelevant. With SOSIG, the user searches a 'focussed' and selective database of resources.

Recording Resources

There is very little meta-data or descriptive information about resources available on the Internet, often no more than a file or directory name - the equivalent of the spine of a book. This can result in users choosing a link or downloading a file, waiting possibly mintues while it transfers to their system (often from the US) only to find it wasn't what they wanted at all. All the resources that appear on SOSIG have been catalogued using a standard pro-forma or template. The template, which includes a description of the resource, underlies the search mechanism which is available on SOSIG. A keyword search will provide you with a list of resources that match your criteria, each of which will dynamically link you to the resource described, wherever it is in the world.

Classification

Each resource is classified using the Universal Decimal Classification (UDC) Scheme. Use of this scheme was agreed with two other national service providers, NISS and BUBL, to allow for collaboration amongst the projects. Although the UDC underlies the organisation of the resources, a strict hierarchical scheme is not enforced, so if a subject has recently become important enough it can find a place on the top menu. Individual resources can also be cross-classified so that they can be found under several different subject areas. The subject categories can be viewed alphabetically (the default) or according to the UDC.

Developments

In 1995 SOSIG received additional funding from the ESRC to allow continuation of the established service for a further two years. Funding from the Electronic Libraries' Programme (eLib) - which has its roots in the Follet Report on IT in Libraries - has also increased the programme of training and awareness of resources for social science teaching and research with the appointment of a Documentation and Training Officer. Workshops running over two days are provided at host sites for research students, academics, support staff and subject librarians, with hands-on sessions built around self-paced on-line and paper materials and a number of presentations and demonstrations. Training materials are available on-line and for downloading, tailoring and use by UK HE institutions free of charge for non-commercial use with appropriate attribution from the SOSIG home page. Evaluation of the training and the general effect and usefulness of subject-based services is also being undertaken as part of the eLib project.

Underlying SOSIG is ROADS (Resource Organisation and Discovery in subject based Services), a two-year collaborative project funded by eLib. ROADS provides the software tools and standards for SOSIG and a number of other subject-based gateways now being built on the SOSIG model, including the templates for recording resources and tools for indexing and searching. The project builds and extends on the work on resource descriptions already in process on SOSIG. The project will also implement a system that allows users to search across several different subject based services seamlessly. This will initially be piloted on the SOSIG service, OMNI (Organising Medical Networked Information) and the Electronic Libraries Information service at the UK Office of Library Networking (UKOLN).

EU funding for the DESIRE project involving eight European partners will build further on the software tools and standards underpinning the gateway and will allow SOSIG to offer a European focus as well as the existing UK and more general 'worldwide' sectors. Consideration will also be given to the extension of cross-database searching, improvements to search mechanisms, meta-data, indexing and cataloguing standards and practices as well as issues raised by the multi-lingual nature of many of the resources.

For more information about the ROADS project contact the authors or see the URL:
<http://ukoln.bath.ac.uk/roads/>

Information about DESIRE is available at:
<http://www.nic.surfnet.nl/surfnet/projects/desire/>

Access to SOSIG
Users with World Wide Web clients such as Netscape or Mosaic can access SOSIG by typing the URL
<http://sosig.ac.uk/>

Users without WWW clients can access the service using the Lynx client. This will give you a text only based interface to the service. Make a telnet call to:

sosig.ac.uk
login as: sosig

Authors:
Debra Hiom, SOSIG/DESIRE Research Officer
Lesly Huxley, SOSIG/DESIRE Training Officer

CHARACTERISTICS OF SOCIAL SCIENCE INFORMATION: A SELECTIVE REVIEW OF THE LITERATURE.* PART I

H. P. HOGEWEG-DE HAART**

Gerrit van der Veenstraat 122, 1077 ER Amsterdam, The Netherlands

ABSTRACT

This paper presents a review of mainly English language literature on the scope of the social sciences as understood in various countries, on the linkings between social science disciplines, on the general characteristics of social science and of social science information, on the characteristics of the primary sources and of secondary information services, on the characteristics of the flow of information, and on the characteristics of the use of social science information by social science researchers as well as by social scientists in non-research environments and non-social scientists in need of social science information, especially in decision-making processes. As far as possible comparisons are made with information in science and the humanities.

WHAT ARE THE SOCIAL SCIENCES?

As the term 'social sciences' is used in various ways in various parts of the world, it seems useful to start an article on the characteristics of social science information with a paragraph on the scope of the social sciences as understood in a number of countries and by some international organizations, with an indication of what the author has in mind when using the term.

In the introduction to the *International encyclopedia of the social sciences* the following comment is given:

> It is apparent that the question 'What are the social sciences?' is one to which no final answer can be given, since—like other groupings of scientific and academic fields—the social sciences differ in their scope

Table 1. Diversity of the social sciences

Columns (left to right):
Yugoslavia*, USSR*, USA‡, UK†, Sweden*, Spain*, Rumania*, Poland*, Norway†, Netherlands†, Italy*, Hungary*, GDR*, France**, Finland*, FRG†, Denmark†, Czechoslovakia*, Bulgaria*, Inform. Services OECD, Dewey Dec. Cl., Intern. Enc. of S.S., UDC, Unesco

Rows (disciplines):
Anthropology (social, cultural)
Ethnography, ethnology
Archaeology, prehistory
Art
Communication science
Criminology, penology
Culture
Customs, folklore
Demography
Economics
Education, pedagogy
Environmental/urban/regional planning
Ergonomics
Futurology
Geography (human, economic, sociological)
History
Information/library science, computing
Labour, occupational sciences
Languages, literature, philology
Law
Linguistics
Management
Military science
Musicology
Philosophy
Political ...

Table 1 (Cont.)

	Public administration	Psychiatry	Psychology	Social policy, social administration	Sociology	Statistics	Science of religion, atheism	Science of science	Scientific communism	Study of developing countries	Trade, traffic
Yugoslavia*					×						
USSR*					×		×	×	×		
USA‡		×	×		×	×	×				
UK†			×	×	×						
Sweden*			×		×						
Spain*					×						
Rumania*					×	×				×	
Poland*	×				×			×			
Norway†			×	×	×						
Netherlands†	×		×		×	×					
Italy*			×		×	×			×		
Hungary*	×		×		×	×	×	×		×	
GDR*					×			×			
France**	×		×		×		×	×			
Finland*				×	×						
FRG†			×	×	×		×				
Denmark†	×		×	×	×	×					
Czechoslovakia*			×		×						
Bulgaria*			×	×	×	×	×	×	×	×	
Inform. Services OECD			×	×	×	×					
Dewey Dec. Cl.	×		×	×	×				×		
Intern. Enc. of S.S.		×	×		×	×					
UDC	×		×	×	×			×			
Unesco		×		×							

from one generation to another. There are also within-generation differences: witness the continuing controversies over whether history should be considered as one of the social sciences or as a humanistic discipline; whether geography is an independent social science or the synthetic discipline that draws upon both the social sciences and the earth sciences; whether law is a social science or a body of professional knowledge; whether psychology belongs with the social or the natural sciences; and whether psychiatry is a social science or a branch of medicine (*International encyclopedia of the social sciences*, 1972: XXI).

Therefore, it may serve better to turn from theory to what internat: classifications and national bodies include in the term 'social sciences practical purposes. For the survey given below, Unesco's list of the sciences was compared with the subdivisions of the UDC, the contents international encyclopedia of the social sciences*, the *Dewey decimal classification*, and classification of the *Inventory of information resources in the social sciences*, com by Brittain and Roberts (1975). These international classifications were compared with what national bodies considered as social sciences.

For each country the papers for the first European Conference on S Science Information and Documentation (ECSSID I), Moscow 1977, scanned, as well as the data collected for the Workshop of ECSSID Worl Group II on ongoing research, Amsterdam 1978. For the USA the tabl contents of the *Social science citation index* was used.

The data in these sources are not quite comparable, as they are not give answers to specific questions. Moreover, the terminology for the discipli included is not always the same; for instance Unesco lists ethnography ethnology next to social and cultural anthropology, but in quite a few count both will be included in the latter. In spite of these incongruities and inco parabilities, however, the listing below gives some insight in the diversity of term 'social sciences'.

It is notable in this survey that there is complete unanimity that econom and sociology belong to the social sciences and almost total unanimity for la Political science and education score for a great majority; history scores well the national context but poorly in international classifications, and opinions the position of psychology are very much divided. The major differenc however, is whether or not the humanities should be included in the soc sciences. Eastern European countries, France and Italy do include humanities, the other countries do not.

Speaking in very general terms the social sciences deal with society and t: humanities deal with the products of the human mind. If the social scienc proper and the humanities are considered as one category it is difficult to spe. of the characteristics of social science information, as it makes a cruci difference whether it is information on the functioning of human society information on art, languages, literature, music or philology. For the purpo of this article the terms 'social science' 'social science information a. confined, therefore, to the more restricted definition.

Another reason to distinguish between social science information an information for the humanities is that interdisciplinary contacts between bou fields are rare, even rarer than between social sciences and technology between social sciences and life sciences.

89

LINKS BETWEEN SOCIAL SCIENCE DISCIPLINES

As to the integration and differentiation processes between social science disciplines it could be argued that, whereas in science, biology, for example, started as one subject and has branched into many disciplines and sub-disciplines, those disciplines linked together under the term 'social sciences', such as anthropology, economics, geography, law, political science, psychology and sociology, each started on its own and links among them developed later. These links not only developed in the way of subdisciplines such as economic sociology or social psychology, but, even more importantly, through interdisciplinary subjects as criminology, education, finance, management, labour sciences, environmental planning and social policy. Of the constituent social sciences mentioned above, some, such as economics, psychology and sociology have a larger input in interdisciplinary subjects than others, such as anthropology and geography.

Links between subjects can be traced by citation patterns. In the DISISS (Design of Information Services in Social Sciences) studies such an analysis was performed. In the conclusion it was stated:

> Subject relationships are shown both by references made (indicating dependence of a subject on other subjects) and by citations received (indicating the extent to which the subject is drawn on by other subjects). These references indicate that there is a heavy general dependence on psychology and to a lesser degree on economics . . . These two subjects are quite self dependent themselves, making and receiving a very high percentage of their references to and from themselves (Bath University, 1979: xiv).

In a more recent article on the same subject Line adds:

> High self-sufficiency can indicate 'inbreeding'—an unwillingness to seek enrichment from other subjects—or maturity and coherence. There would be little disagreement that psychology and economics are the social sciences with the clearest identities. Sociology and political science are of their nature less clear cut, while the dependence of geography, criminology, and of education on other subjects is understandable (Line, 1981: 81).

James D. Neeley (1981: 222) confirmed the usefulness of the method of cross-citation analysis in a recent article on the interdisciplinary position of the management literature.

Linkages among sciences have been distinguished as:

1. Intradependent, when its research is observed to feed more on the research produced by itself than on the research produced by the other science.
2. Interdependent, when the research on one is observed to feed on the research of another in equal proportions.
3. Homodependent, when its research is observed to feed on the research of its assumed parent-science.
4. Heterodependent, when it feeds more on exterior sciences than on itself and related sciences (Espirito Santo, 1978: iii).

These linkages among subjects may be expected to have an important impact on the use of the literature for these subjects.

GENERAL CHARACTERISTICS OF SOCIAL SCIENCE AND () SOCIAL SCIENCE INFORMATION

In 1974 Unesco sponsored a meeting of social science experts to investigate feasibility of a world information system for the social sciences in the fr. work of the UNISIST programme. Adam has summarized the characteri of social science information as brought forward in the report of this Vales meeting. In an abridged form they are:

1. Materials in the social sciences are needed by a much wider range of users. User groups include not only scientists but also administrators, policy makers, businessmen, social workers, lawyers, etc.
2. Conversely, in most of the social sciences, the range of potential subject matter is extremely wide.
3. Certain types of information and data that are important in the social sciences are not found in other fields, or are of much less consequence there, as for instance: conceptual information, statistical and other numerical data, data which are initiated in non-Roman characters.
4. Much information is of restricted value in terms of time and/or place.
5. Social science knowledge can be packaged in a wide variety of ways, e.g., by subject matter, problem, geographical area, professional need.
6. Special forms of output are often necessary to satisfy the needs.
7. Social science materials are recorded in a wide range of media and formats. The organization and description of such varied formats present problems when related to users' needs.
8. Social scientists place a heavy reliance on informal means of communication and are often reluctant to use formal channels.
9. There is a great deal of reluctance to use data at a secondary level or to replicate earlier research.
10. The boundaries between disciplines are neither static nor clearly defined and the scope of social sciences as a whole is subject to varying definitions.
11. Social scientists' work is more significantly affected by their own world views which are influenced by social, national, religious, linguistic, cultural and ethnic differences (Adam, 1975: 287)

Some years later Brittain set the characteristics of science informatio systems and the structure of communications and knowledge against those i the social sciences. Some of his conclusions are:

1. In the social sciences there is less agreement about the way in which progress is made, the relative importance of communications channels, and information requirements.
2. Progress in the social sciences does not seem to proceed by building upon the achievements of previous generations.
3. There is no general consensus about subject matter, procedures, methods, and interpretation of data.
4. The social sciences suffer from a wealth of theories that either cannot be verified by empirical evidence, or are not verified because social scientists do not persist when there is conflicting evidence, which leads to rival schools and to a mass of literature with a lack of consensus.
5. The lack of consensus and problems of verification mean that many

91

theories amount to no more than individual points of view (Brittain, 1979: 713).

In Brittain's opinion, to adjust information services to the requirements of the social sciences, it is necessary to study more deeply the differences in information flow and in communication and the structure of knowledge. In particular he suggests that studies by means of citation analysis should be made of publications in some ten areas of the social sciences which are generally recognized to deal with important issues.

One of the points Brittain stresses strongly is that for the most part the social sciences are non-cumulative.

> This contradicts the expressed views and hopes of many social scientists over the past 100 years; and also information specialists, who often assume that social science information accumulates to form a structured body of knowledge (Brittain, 1979: 723).

CHARACTERISTICS OF THE SOURCES OF SOCIAL SCIENCE INFORMATION

Characteristics of the primary literature

In this section data will be given on the size and composition and of the growth rates of social science literature. Wherever possible these data will be compared with data on literature in science and the humanities.

The data for this section, except where indicated, are drawn from the publications on the DISISS project (Bath University, 1974, 1975, 1979, 1980). For the purpose of this project the scope of the social sciences was not limited to the definition given above but included also geography, linguistics, social and economic history and statistics, and excluded law.

Size and composition of social science literature

According to the DISISS team there is a good deal of literature discussing and reporting size and growth data, at the global, national and subject level, but the comparability among the studies is low. Estimates on the world number of social science serials vary from 2662 serials in the 1973 World List of Social Science Periodicals (scientific primary periodicals only) to 27 000 serials in the Unesco Statistical Yearbook 1972. This last number is thought to contain much non-academic material, and a better estimate might be in the order of 7000.

Line and Roberts (1976: 126) suppose that approximately 20 per cent of all 'serious' current serials might be social science serials. They estimate a production of 140 000 social science articles in 1970 against 130 000 monographs, which makes for a rate of 1.08:1, whereas this rate is 8:1 in science and technology.

For the bibliometric studies of the DISISS project a machine-readable database of information on social science serial titles, named CLOSSS (Check List of Social Science Serials), was constructed, containing 3909 current titles, 728 previous titles and 1595 discontinued titles. The CLOSSS file is not comprehensive. Non-English language titles are certainly under-represented

and an under-recording of serials prior to 1950 is suspected by the DIS team. However, it seems to be the most serious study available.

The subject distribution of the 2760 current social science serials i: 8.55 per cent social sciences general, 2.53 per cent anthropology, 0.83 per archaeology, 0.50 per cent architecture, 1.12 per cent criminology, 25.76 cent economics, 10.36 per cent education, 2.39 per cent environme cent planning, 0.28 per cent ergonomics, 0.18 per cent futurology, 4.87 per geography, 2.10 per cent history, 4.89 per cent law, 2.17 per cent librar ship, 5.14 per cent linguistics, 2.60 per cent management, 0.50 per philosophy, 9.97 per cent political science, 6.41 per cent psychology, 4.74 cent social work, 3.76 per cent sociology and 1.05 per cent statistics. T figures make clear how prominent economic literature stands among the s science literature on the whole, followed by the literature on education an political science.

In contrast with the data for social science serials which could be drawn fr the CLOSSS file, the data for social science monographs had to be produced analyses from other statistical data. The DISISS team makes an estimation 750 000 items of all types of monographs together, produced annually. T conclude that:)

> The monograph literature of the social sciences appears on a crude estimate to be slightly smaller than the humanities literature (including modern literature and fiction) and the combined pure and applied science literature (Bath University, 1974: 76).

However, the social sciences (classes 3 and 9 *Dewey decimal classificatic* produced some 9000 titles (10.1 per cent) more than the pure and appli sciences together in 1970, which implies that:

> Social science literature has increased to a level at least comparable with other broad areas, and at a rate which suggests it will soon represent the largest area (Bath University, 1974: 77).

As to the subject distribution of monograph literature, 28.6 per cent wer, classed (according to Unesco's book production statistics) as politics an economics, 15.6 per cent as education, 14.5 per cent as biography and histor and 9 per cent as commerce. Twice as many titles were produced in educatio: and four times as many in politics and economics, compared with sociology.

Growth rates of the literature

The growth rate of social science serials in the period 1950–1970, as analyse from the CLOSSS file, is 3.35 per cent p.a. for the world, with a rather larg variance between countries. Growth in single subject areas of the socia sciences varied widely. In the period 1960–1970 economics and education wer the faster growing disciplines, while sociology, political science and psycholog had a slower growth rate.

, Citation analyses indicate that references in serials show a much faster deca than references in monographs in nearly all subjects and for all forms o material cited. Among subjects economics and political science show the faste decay rates.

Also for monographs the growth rate differs considerably between soci science disciplines, being the lowest in sociology and statistics, followed b

philosophy, psychology, trade and commerce, education, geography, and history, the highest being politics and economics.

Contrary to the growth rates in the social science monographs, the number of monographs in science and technology show a very low level of growth and the growth of the number of monographs in the humanities is even lower (Line and Roberts, 1976: 33).

As to the future: owing to the economic crisis on the one hand and technological change on the other, a prediction of future growth rates would be very hazardous:

> In the short term, a decline in growth rate may ease the problems of bibliographic control, but any benefits here will be more than offset by the likely increase in semipublished material (report literature and working papers); in non-conventional publications such as microforms and synopsis journals, and in on-demand publishing (Bath University, 1980: 22).

The role of the electronic journal in social science information seems to be quite uncertain.

Other primary sources

Whereas the library is sometimes called the social scientist's laboratory, his ingredients are not only the library's holdings of serials and monographs, but also other primary materials that are more often than not insufficiently available in libraries.

The range of materials that is potentially of use to social science research was enumerated by J. Madge in *The tools of social science* (London, 1953), cited by Brittain. They include:

1. Personal documents, including letters, folklore, life histories, autobiographies, diaries and letters.
2. Records, including records of professional societies, committee records, government records.
3. Reports that are made after an event, e.g., newspaper reports.
4. Observations, including mass observation where the investigator may penetrate into the environment he is observing.
5. Data from 'action research' where the investigator resides in the target area for a period to observe the group life, morale and productivity of a single community with the aim of developing effective ways of resolving social stress and tension and facilitating agreed and desired social change.
6. 'Overheard' data from passive observation.
7. Data from interviews such as Gallup polls, opinion polls, scalogram analyses, and measures of attitudes.
8. Data from experiments, usually conducted in controlled environments in laboratories (Brittain, 1970: 36, 37).

P. L. Garvin in *Problems of processing information in the behavioral sciences*, an unpublished paper (1967), also cited in Brittain, adds to this enumeration the non-textual data, especially used in social psychology and psychiatry, such as gestural movements, facial expressions and psychiatric interviews (Brittain, 1970: 37).

Line (1971: 203), in his conclusions on the Aslib Social Sciences Group's Conference on primary materials in 1971, mentioned some uncertainty as to

94

the clear division between primary and secondary material. As examples of primary materials he gave: unpublished survey reports, raw statistics, personal reminiscences, records and minutes of meetings, local political manifests, parish records, etc., and also non-book material, including sound recordings, videotapes, pictorial media such as air photographs and machine-readable files.

During the Aslib Social Sciences Group's Conference, papers were given on primary materials in economics, in politics and political science, and in urban and regional planning.

Swan (1971: 167) mentioned as the main forms of primary materials for economists in the UK those which emanate from the UK government, the international bodies, such as the UN, OECD, EEG, EFTA, IMF and foreign governments.

As primary materials used in politics and political science Rush (1971: 178–181) lists:

1. Original documents, including personal documents, formal records and reports (unauthorized accounts and comments).
2. Digests, annuals, yearbooks.
3. Opinion polls and other surveys.

White (1971: 191–196) lists as primary sources most needed by urban and regional planners:

1. Topographical and thematic maps and plans.
2. Data collected from national and local surveys.

The principal gaps in which planners require more and better data she considered to be: income, population migration, and interregional movement generally. The main problem in using data is the compatibility among data drawn from different sources.

These examples of the need for primary materials in three social science disciplines make it clear that although these primary materials are a very characteristic feature of social science information they are a hardly explored area. Therefore, details will be given in only three of the better studied categories of primary materials: ephemera, statistics, and machine-readable data.

Ephemera

According to Pemberton ephemera may be defined as:

> documents which have been produced in connection with a particular event or item of current interest and which are not intended to survive the topicality of their message (Pemberton, 1971: 6).

They are the raw material for much of the research undertaken by social scientists and social historians. Pemberton (1971: 53–54) lists 39 types within this definition, including advertising circulars, company reports, computer programs, conference papers, manifestos, newsletters, newspaper cuttings, obituaries, photographs, press releases, questionnaires, survey reports, transcripts of broadcasts, annual reports, broadsheets, constitutions and rules, leaflets, membership lists, pamphlets and posters.

Pemberton carried out an investigation of collections of printed ephemera in national libraries, university libraries, libraries of research institutes, government libraries, specialist libraries and public libraries in the UK. He studied acquisition policies and views on a central collection of printed ephemera. He found rich collections in every category of library. For instance:

1. Public libraries often had rich collections in local history.
2. Newspaper, radio and television services were rich in all forms of topical ephemera.
3. Political party libraries had much historical and topical material.

But the collections were scattered and there was no guide to them. Demand was high, in many cases even too high to handle adequately. Problems were considered to be:

1. Difficulties with identifying, obtaining and processing material of this kind.
2. The impossibility of bibliographic control because of a lack of bibliographies.
3. The difficulty of forecasting which trends or movements of today are going to be the stuff of research tomorrow (Pemberton, 1971: 45).

Pemberton made a strong case for a central collection of printed ephemera for which he listed a considerable number of advantages. But the first thing needed would be a directory to the existing collections. Moreover:

> Any systematic approach to the provision and preservation of ephemera on a national scale must begin with a survey in depth of the needs of the users, and continue by modifying policies in response to changing needs revealed by constant surveillance (Pemberton, 1971: 45).

Statistics

Statistics are also a primary source characteristic for social science research. They are used and produced in great quantities, especially in economic research, and even more in all levels of national, regional and local administration. However, many statistics are collected in different categories and time scales, which makes it difficult to compare them, not only on an international scale, but even on a national, regional or local scale. This difficulty is well illustrated by the library statistics as given earlier in this article.

Piatier suggests that:

> A world statistical centre should be able to deal with all the statistical material available, make series homogeneous, explain divergencies, and indicate the degree of reliability (Piatier, 1976: 440).

On the other hand, as statistics are collected not only by governmental, but also by professional and private organizations, on multinational, national and local scales, there would seem to be a useful place for a general directory of possible sources of information. To parallel other types of scientific information the following characteristics should be included:

> periodicity or field covered, time required to obtain information, length of the series, number of countries for which data are comparable (Piatier, 1976: 441).

96

So far as to published statistics. A matter of concern should also be the 'concealed statistics', i.e., those statistics calculated by research workers on the basis of official and other published statistics, but not included in organized information systems and networks (Földi, personal observation).

Machine-readable data files

Machine-readable data files are another primary source for social science research. Robbin (1981: 97–109) draws attention to the problems in the utilization of numerical databases. She argues that the large databases which are developed to respond to administrative reporting requirements and used for research and policy questions must be organized in tractable ways, and she mentions a number of strategies to improve the unsatisfactory situation.

Machine-readable files of concluded research projects are collected by specialized social science data archives for the benefit of social science research, as the raw data of one researcher's project may, by secondary analysis, enter as a primary source into a second researcher's new project. These data archives make the data more easily accessible through documentation and the production of, for instance, SPSS files. Nevertheless, the data are not used as much as they could be in longitudinal or comparative research.

One of the reasons for non-use, suggested by White, is that the data files are, just like ephemera, not entered into the secondary information systems. The reason for this is that the unit for bibliographic description of these files differs from that of printed material. White argues that data files have two parts: the codebook and the machine-readable part. In file description the codebook should be a primary source of the abstract, it should be subject-indexed and it can serve for library cataloguing. According to White, another way of improving the accessibility of machine-readable data files is to put abstracts of data files, and instructions for ordering their codebooks, into an existing national online bibliographic service, as NTIS does for data files made public by government agencies (White, 1977: 313–322).

Dodd (1979: 77–82) proposes that a proper bibliographic reference for these files should include: authorship, title, subtitle, general material designator, statement of authorship (including the sponsor or funding source), edition, imprint (producer statement and distributor statement), notes and series. A bibliographic citation, incorporating all the required and optional fields of a main entry, is recommended in the *Anglo-American cataloguing rules* (American Library Association, 1979). One of the optional elements is the medium designator MRDF, which stands for Machine Readable Data File (Robbin, 1981: 102, 105).

CHARACTERISTICS OF SECONDARY INFORMATION

Indexing and abstracting services

In the overview of research carried out for the DISISS project, it is stated:

One major difference between science and technology on the one hand and the social sciences on the other is the nature of the potential market for secondary services. The social science market is smaller in total, more diverse and less easy to identify, and above all much less wealthy (Bath University, 1980: 85).

This corresponds with Clayton's statement:

> In science there is a small number of very large services, well financed
> with smaller overlapping services alongside. In the social sciences and
> the humanities there is a very large number of small secondary services
> (Clayton, 1976b: 14).

Brittain and Roberts (1975) listed 295 organizations in Western Europe
which provided social science information services: 7 for anthropology, 9 for
criminology, 3 for demography, 52 for economics, 46 for education, 17 for
environmental planning, 1 for ergonomics, 1 for futurology, 3 for geography, 4
for history, 1 for linguistics, 15 for management, 30 for political science, 6 for
psychology, 26 for social policy and social administration, 11 for sociology,
8 for statistics, and 54 for social and behavioural sciences in general. Never-
theless:

> As yet bibliographical control in the social sciences is, both relatively and
> absolutely underdeveloped, a situation which should not long be allowed
> to continue (Line and Roberts, 1976: 77).

In an International Workshop of Secondary Service Producers held in York
in 1975, it became evident that the key problem is that although the number of
secondary services is great (980 in 1970 of which 200 were considered to be
major) (Line, 1976a: 7) and the rate of growth of the services has since about
1915 been greater than the rate of growth of the primary literature (Brittain,
1976: 19) usage is certainly less than it might be, particularly by practitioners
(Clayton, 1976a: 5).

> Most secondary services are very deficient in covering non-serial
> literature, an important matter, since whereas in science non-journal
> items account for only about 10% of the literature, in the social sciences
> they account for about half (Line, 1976a: 7).

Freides (1976: 70) presented data on coverage of books and US government
publications by the bibliographic services of psychology, political science and
economics and two interdisciplinary bibliographies. Coverage of books ranged
between 45 per cent and 68 per cent; coverage of government publications was
about 50 per cent. In accordance with the DISISS findings coverage increases
with the number of bibliographic services, but so does duplication.

The DISISS team paid much attention to the clustering of journals from a
file of nearly 50 000 citations drawn from a representative sample of journal
articles in most of the social sciences. Clustering of journals and a comparison
of clusters with existing coverage patterns gives a good insight into the value of
specific services for user groups. This could be an incentive for rationalization.
Line (1976b: 53) suggests it would be beneficial to get secondary service
producers to define clearly their aims and their audiences.

A possible future system is suggested by the DISISS team which in principle
would consist of:

> An international database of references maintained in high density
> storage, accessible world-wide for both input and use. New items would
> be fed into the database, which would first be checked to avoid unneces-
> sary duplication; references would not be duplicated, but additional
> index entries and different abstracts might be entered to reflect different
> perspectives (Bath University, 1980: 87).

However, as the obstacles to such an 'ideal' system seem formidable, moves towards compatibility between the already existing machine-readable social science databases might benefit the whole system in such a way that total integration would not be necessary.

Review articles and book reviews

There are few pertinent data on review articles and book reviews in the social sciences. The DISISS team reached the following conclusions:

> Altogether in 1972 the output of scientific reviews of reasonable quality (according to criteria identified by Woodward) was about 22 000, which is rather less than 2% of the world's scientific and technical literature . . . An estimate of the number of reviews per year in the social sciences has not been produced, although a rough guess is possible . . . Allowing for growth in both journals and review articles (which is now quite likely) it is possible that in 1974 about 1500 review articles were produced in social science journals . . . However rough the calculation, it is clear that the ratio of reviews to primary articles is much smaller in social science than in science: the ratio of reviews in science to reviews in social science was about 16:1 in 1972; the ratio of reviews to primary articles is approximately 1:45 in science and technology, and 1:133 in the social sciences in 1972 (Bath University, 1975: 114, 115).

Line (1976b: 52) remarks that there are also reviews 'concealed' in monographs but does not give an estimate of their number in relation to reviews in serials.

Also in the case of reviews there are considerable differences between the various social science disciplines. Of the CLOSSS file 41 per cent of the linguistic serials, and 27 per cent of education serials contained review articles, whereas in social welfare, criminology and economics the proportion of serials containing review articles was around 5 per cent.

As to the humanities the DISISS team concludes:

> Critical book reviews are important in nearly all fields, but there is no evidence to suggest that quantitatively or qualitatively they are more important in the humanities than in the social or natural sciences (Bath University, 1975: 116).

CHARACTERISTICS OF THE FLOW OF INFORMATION

The American Psychological Association's Project on Scientific Information Exchange in Psychology (APA/PSIEP) is a classic study on this subject, executed in the early sixties. The Project's first objective was to develop a natural history of scientific information exchange, i.e., to describe the scientific exchange environment of the scientific psychologist. Most of the earlier studies of the Project dealt with each of the channels of communication in psychology, such as journals, conventions, the distribution and use of reports prior to publication, and the informal face-to-face communication (APA/PSIEP I, 1963).

The research directors Garvey and Griffith consider after analysing these earlier studies on the process of dissemination and the behaviour of the active user that:

> It seems clear that while information retrieval services mainly wait for 'public' information (i.e., from or in archival sources), the scientist who wants contemporary findings to plan research or to interpret his own findings does not. During the various stages leading to journal publication of work, he is involved in trying to discover every means of obtaining information on new, ongoing, or recently completed work relevant to his own. He does not seem willing to wait to discover this in a journal or a secondary source; rather he seems to use journals to catch what he has missed in his efforts to gather information in the past couple of years (APA/PSIEP I, 1963: 10).

Later findings of the Project have hinted that scientific communication and information exchange cannot simply be regarded as a single system serving the same needs and transmitting the same type of information to all users. More research was required on the information need of applied psychologists, on the ways in which these needs are met, and on the operation and use by applied psychologists of media outside the information system. More research was also required on the informal communication and social organization among active researchers, and on national environments as a factor in the development and functioning of psychology (APA/PSIEP III, 1969).

Garvey, Lin and Nelson of the Johns Hopkins University's Centre for Research in Scientific Communication compared the processes of disseminating and assimilating information in science and in social science. In this study differences were ascertained between time-lags in the scientific social scientific information flows. For instance:

1. The delay between the last prepublication report and the time of journal publication is nine months for scientists and 15 months for social scientists.
2. Prepublication dissemination of information appeared less effective in the social sciences than in the sciences.
3. More physical scientists reported their work prior to a national meeting than social scientists.
4. Social scientists show greater activity in and resulting from meetings than physical scientists.

The authors conclude that, although as a whole the scientific communication process shows gross similarities, there are considerable differences between disciplines, and communication innovations designed for one discipline may prove inappropriate and even damaging for another (Garvey et al., 1970: 1166–1173).

Brittain (1979) makes a comparison between the structure of communication and knowledge in science and the social sciences. He argues that because of the differences found in these structures the information services for the social sciences, which have hitherto largely been based upon the model of science information services, should be adapted to the requirements of the social sciences. He indicates the studies necessary to establish these adaptations and he stresses the growing role of librarians and information specialists in the process.

Another approach to analyse the flow of information in social science is citation analysis. According to the DISISS team:

> The chief interest of citation studies lies not in practical library applications, but in potential application to information system design and

150

interpretation for studies of the communication and transmission of knowledge (Bath University, 1979: 3).

CONCLUSIONS

1. There is no unanimity on what is meant by the term 'social sciences The most striking difference is that some countries include the humaniti in the social sciences and others do not. As the object of social scien in its proper meaning is crudely 'man in relation to society' and as an and letters are crudely spoken 'the spiritual products of man', inter disciplinary contacts are rare.

2. Some social science disciplines show more linkings to each other th others. One can speak of self sufficient and dependent disciplines, di tinguishing between various types of dependency. Psychology, economi and to a lesser degree sociology are more self dependent and are cite extensively by other disciplines and interdisciplinary subjects. Therefore their literature will be used far beyond the scope of their own field.

3. The characteristics of the social sciences are reflected in the chara teristics of social science information, i.e., the often unclear and eve changing language of the social sciences, the different meanings of term as used in various subjects, and the considerable overlap betweer disciplines are reflected respectively in the incompatibilities of informa tion over time sequences, the terminological problems of social scienc information retrieval, and the overlap between information services.

4. The sources used for this review are almost exclusively taken from th English language literature and research carried out in the UK and USA Little or none could be found in the French or German or Dutch literature, and it does not seem likely that the characteristics of socia science information as such constitute an important subject of research in other Western European countries or in Eastern Europe.

5. The production of social science journal articles is hardly higher than the production of monographs, whereas in science the number of articles is eight times as large as the number of monographs. The monograph literature in social science is estimated to be only slightly smaller than for either the literature of the sciences or the humanities. Growth rates of the social science literature are greater than those of the literature of the sciences or the humanities. Therefore, the amount of primary literature social science information will have to deal with in the near future will be comparable with or larger than the quantity of information which has to be processed in science or the humanities.

6. Apart from serial and monograph literature other primary materials are very important for social science research. These primary materials are insufficiently available in libraries, and information on this kind of material is for several reasons quite unsatisfactory, especially for ephemera. Handling statistical material and machine-readable data in a way comparable to other bibliographic rules would make them more easily accessible for the user.

7. Secondary information services in the social sciences are small and numerous. Bibliographic control is unsatisfactory; there are gaps and overlaps in coverage. Provision should be made for different categories of users.

101

8. The flow of information between social scientists is slower and more laborious than between scientists. Possibly as a result of the difficulties in formal communication, social scientists seem to show in some respects a greater activity in informal communication, though research results are somewhat contradictory.

The implications of these conclusions for information services in the social sciences would appear to be the need for more cooperation in order:

1. To enhance the coverage of all primary literature, including primary materials other than books and serials.
2. To enhance compatibility among databases.
3. To avoid unnecessary overlap.
4. To make a common effort towards a greater compatibility among retrieval languages.

REFERENCES

ADAM, R. (1975). A world information system for the social sciences. *Aslib Proceedings*, 27, 286–293.

AMERICAN PSYCHOLOGICAL ASSOCIATION. *Project on Scientific Exchange in Psychology Reports.* Washington D.C., Vol. I (1963), Vol. II (1965), Vol. III (1969).

BATH UNIVERSITY. (1974). *Size and growth of monograph literature with particular reference to the social sciences.* Bath: Bath University, Bath University Library. (DISISS Working Paper No. 7)

BATH UNIVERSITY. (1975). *Size, growth and composition of social science literature.* Bath: Bath University, Bath University Library. (DISISS Research Report, Series A No. 2)

BATH UNIVERSITY. (1979). *The structure of social science literature as shown by citations.* Bath: Bath University, Bath University Library. (DISISS Research Report, Series A No. 3)

BATH UNIVERSITY. (1980). *Towards the improvement of social science information systems; overview of research carried out 1971–1975.* Bath: Bath University, Bath University Library. (DISISS Research Report, Series A No. 1)

BRITTAIN, J. M. (1970). *Information and its users: a review with special reference to the social sciences.* Bath: Bath University Press.

BRITTAIN, J. M. (1976). Relationships between secondary services: coverage, overlap and scope for co-ordination. In: *International workshop of secondary service producers*, pp. 19–37. (BLR&D Report No. 5289)

BRITTAIN, J. M. (1979). Information services and the structure of knowledge in the social sciences. *International Social Science Journal*, 31, 711–728.

BRITTAIN, J. M. and ROBERTS, S. A. (1975). *Inventory of information resources in the social sciences.* O.E.C.D., Farnborough: Saxon House Ltd.

CLAYTON, K. M. (1976a). Introduction to the international workshop. In: *International workshop of secondary service producers*, p. 5. (BLR&D Report No. 5289)

CLAYTON, K. M. (1976b). Discussion. In: *International workshop of secondary service producers*, pp. 14–15. (BLR&D Report No. 5289)

DODD, S. A. (1979). Bibliographic references for numeric social science data files: suggested guidelines. *Journal of the American Society for Information Science*, 30, 77–82.

ESPIRITO SANTO, A. DO (1978). *A measure of the dimensions of interdisciplinarity of two applied sciences: a scientometric model.* Madison, Wisc.: University of Wisconsin (thesis).

FREIDES, T. (1976). Bibliographic gaps in the social science literature. *Special Libraries*, 67, 68–75.

GARVEY, W. D., LIN, N. and NELSON, C. E. (1970). Communication in the physical and the social sciences: the processes of disseminating and assimilating information differ in these two groups of sciences. *Science*, 170, 1166–1173.

102

INTERNATIONAL ENCYCLOPEDIA OF THE SOCIAL SCIENCES (1972). New York/London. Macmillan.

LINE, M. B. (1971). Primary materials in the social sciences: conclusions. *Aslib Proceedings*, *23*, 203-206.

LINE, M. B (1976a). Research into information systems design. In: *International workshop of secondary service producers*, pp. 6-8. (BLR&D Report No. 5289)

LINE, M. B (1976b). Concluding summary. In: *International workshop of secondary service producers*, pp. 52-54. (BLR&D Report No. 5289)

LINE, M. B. (1981). The structure of social science literature as shown by a large-scale citation analysis. *Social Science Information Studies*. *1*, 67-87.

LINE, M. B. and ROBERTS, S. A. (1976). The size, growth and composition of social science literature. *International Social Science Journal*. *28*. 122-159.

NEELEY, J. D. Jr (1981). The management and social science literature: an interdisciplinary cross-citation analysis. *Journal of the American Society for Information Science*. *32*. 217-223.

PEMBERTON, J. E. (1971). *The national provision of printed ephemera in the social science.* Warwick: University of Warwick Library. (Occasional Publications No. 1)

PLATIER, A. (1976). General reflections on information and international statistics. *International Social Science Journal*, *28*, 425-448.

ROBBIN, A. (1981). Strategies for improving utilization of computerized statistical data by the social scientific community. *Social Science Information Studies*. *1*, 89-109.

RUSH, M. (1971). Primary materials in politics and political science. *Aslib Proceedings*. *23*. 175-186.

SWAN, D. (1971). Primary sources in economics. *Aslib Proceedings*, *23*, 167-174.

WHITE, B. (1971). Primary materials in urban and regional planning. *Aslib Proceedings*, *23*, 187-198.

WHITE, H. D. (1977). Toward a national information system for social science data files. *Journal of the American Society for Information Science*, *28*, 313-322.

103

The survival of the fittest is quickly becoming the survival of the fittest to learn. Unless an organisation continuously adapts to the environment via speedy, effective learning, it will die (Harrison Owen, 1991).

Definition: learning organization and living company

The quote below gives a simple definition of a learning organisation :

> 'A learning organisation is one in which people at all levels, individuals and collectively, are continually increasing their capacity to produce results they really care about.'

Dr Peter Senge (1998) says that an organisation that encourages continuous learning among its people, promotes exchange of information and knowledge between employees creates a more knowledgeable workforce. This produces a very flexible organization where people will innovate, accept, and adapt to new ideas and changes through a shared vision. Learning organizations balance the individual's and the organization's development needs in order to properly address both.

"New mantra – Learning" - emergence of learning organizations

The learning organization concept is becoming an increasingly widespread philosophy in modern companies, from the largest multinationals to the smallest ventures. The concept of organization-wide learning, and recognition of its importance can be traced back in the research literature back to the 1940s, but since the last two decades organizations began to realize its potential for increasing performance, competitiveness, and success with this "new mantra – learning". The ever-increasing changes of the 21st century have made learning organizations ever more important.

Learning organizations are living companies

They recognize that empowered and enabled employees are essential for success. A continuous improvement culture is clearly a learning culture. The learning organizations collect, create, store, transfer, and use knowledge effectively and productively using IT network systems.

Organizations such as Bechtel, Flour Daniel, Dupont, Wal-Mart, Proctor & Gamble, Cray, MeKinsey, Ernest and Young, FedEx, Ford Motors, Motorola, Knight-Rider, Shell Oil, General Electric, Johnsonville Roods, Quad Graphics and Pacific Bell in the United States; Rover Automobiles, Whirpool, British Petroleum, Sheerness Steel, Nokia, Sun Alliance, and ABB in Europe; and Honda, Samsung in Japan, Tata Steel, L&T, Godrej, Infosys, Wipro, TCS in India are excellent examples of learning organizations.

What is a learning orgnisation?

In general, learning organizations will posses such 14 characteristics as those given below.

1. The organizational system learns as a whole, almost as if the organization were a single brain.
2. People in the organization recognize that ongoing, organization-wide learning is critical for the organization's current as well as future success.
3. Learning is a continuous, strategically used process that is integrated with and parallel to work.
4. Focus is on creativity and generative learning.
5. Belief is that system thinking is fundamental.
6. People have continuous access to information and data resources important to the company's success.
7. The organizational climate encourages, rewards, and accelerates individual and group learning.

104

8   Human resources are innovative and communal in their networking, both inside and outside the organization.

9   The organization embraces change, and views unexpected surprises and failures as opportunity to learn.

10   The organization is agile and flexible.

11   People in the organization are driven by a desire for quality and continuous improvement.

12   Aspiration, reflection, and conceptualization characterize most organizational activities.

13   Well-developed core competencies serve as the launching pads for new products and services.

14   The organization can continuously adapt, renew, and revitalize itself in response to changing environments.

System model

Dr. Peter Senge, in his famous book, *The Fifth Discipline* (1987) has identified six skills that affect learning in an organization. Organization is a system with five interrelated components: the people, the organization, the learning, the knowledge, and the technology. All the five are necessary to sustain viable, ongoing organizational learning and ensuing corporate success.

- *Personal mastery* proficiency in a subject or skill.
- *Mental models* deeply ingrained assumptions that influence how we understand and take action in the world.
- *Shared vision* finding common ground in an organization that fosters real commitment, rather than simple compliance.
- *Team learning* the process of developing a team and learns and produces.
- *System thinking* a conceptual framework that makes patterns (and ways to change them) more understandable.
- *Dialogue* the free and creative exploration of subtle issues, a deep listening to others.

Winning vision and organizational learning

The vision, culture, strategy, and structure of the organization play a key role and shared vision gives individuals and the organization "stars to steer by".

We are all living in a world where all things that used to give business a competitive advantage are disappearing...... The only thing that sets us apart, perhaps the only sustainable competitive advantage we have in an organization – is the ability of our people to learn and act faster than the competition. Successful leaders influence organizational members through the power of the vision – organizational learning which they passionately believe in. In today's scenario organizations need to win even to stay up the race. A winning organizational learning vision creates excitement, captures the imagination of people, provides them a sense of belonging and purpose to move collective energies in the needed direction, helps build strategy, and clarifies the action priorities for building the organizational road map.

Knowledge integration through organizational learning

Dr Peter Senge (1998) – Orgnizational learning and knowledge integration is a powerful competitive weapon in high-performance organizations. Accelerating change is transforming our world. Organizations today face adaptive challenges. Change is inevitable in all aspects of life a but successful person or organization is one who holds on to the old as long as it is good and grabs the new as soon as it is better. Although modern management has developed sophisticated analysis and decision-making tools like quality management, six sigma and benchmarking to help organizations manage change and competition better, these tools are

essentially how to do tools. The solutions to adaptive challenges reside in continually enhancing the core competencies and competitive intelligence of employees through organization learning at all levels.

Dr Peter F. Drucker (2001), one of the top management guru emphasized efficiency, agility, and productivity as today's competitive advantages. The efficient, agile, highly productive organizations do something fundamentally different from others to thrive even in tough economic times. What is the secret? They all apply "knowledge integration' - through organizational learning" as a fundamental management practice throughout their operations. They also consistently execute their missions by focusing on operational excellence (benchmarking) and by continually implementing new (Kaizens), best practices through out the organization. Regardless of industry, location, size or mission, high-performers actively leverage the know-how of their employees – at all levels of the organization – to get things done. They also understand that it's not what you know, it's what you do with what you know that counts and "know what they're doing" – literally!..

Thomas Stewart (1977) writes *In Intellectual Capital: The New Wealth of Organizations* "Knowledge has become more important for organizations than financial resources, market position, technology or any other company asset."

Knowledge is the food of the learning organization, its nutrients enable the organization to grow. Individuals may come and go, but if its valued knowledge is lost, the company will starve to death (Davenport and Prusak, 1998).

Nonaka and Takeuchi (1995) proclaim that an organization's ability to create, store, and disseminate knowledge is absolutely crucial for staying ahead of the competition in areas of quality, speed, innovation, and price. Successful companies are those that consistently create new knowledge, disseminate it widely throughout the organization, and quickly embody it in new technologies and product. Knowledge creation and organizational learning should be at the epicentre of a company's corporate strategy". Only by organizational learning, the fruits of its thinking will a company be able to transform knowledge into corporate power.

Stewart (1997) notes that systematic application of "intellectual capital creates growth in shareholder's value". This is accomplished through continuous recycling and creative utilization of the organization's rich knowledge and experience. Technology enables optimum application of corporate knowledge.

Brain Quinn (1992) author of *The Intelligent Orgnisation,* calls technology the most important ingredient for managing organizational learning and knowledge. Effective use of information communication technology to harness and enhance learning have a significant competitive advantage over those use stone age tools.

Becoming a learning organization

There is no single, guaranteed way to become a learning organization. Each organization must look at its own history, competitive environment, skill base, technology, mission, and culture and then develop an appropriate learning structure and style.

Sixteen steps to becoming a learning organization

1 Commit to becoming learning organization.
2 Connect learning with business operations.
3 Access the organization's capability on all five learning organization systems: people, technology, knowledge, learning, and the organization itself.
4 Communicate the vision of a learning organization.
5 Recognize the importance of systems thinking and action.
6 Demonstrate and model a commitment to learning.

7 Transform the organizational culture to one of continuous learning and improvement.
8 Establish corporate-wide strategies for learning.
9 Cut bureaucracy and streamline the structure.
10 Empower and enable employees.
11 Extend organizational learning to entire business chain.
12 Capture learning and release knowledge.
13 Acquire and apply the best of technology to the best of learning.
14 Encourage, expect, and enhance learning at the individual, group, and organizational level.
15 Learn more about learning organizations.
16 Continuously adapt, improve, and learn.

Competitive advantage

In today's highly competitive business environment, learning organizations hold a significant competitive advantage. Their ability to harness the power of learning at all levels – individual, team, and organizational – enables them to rapidly leverage new knowledge into new products and services, new marketing strategies, and new ways of leading the learning revolution. Orgnizations with big brains and the ability to learn quickly will become global leaders.

Acquire, adapt, and advance

The pace of change is so rapid today and in the high-velocity markets that no single organization can ever control all effective practices and good ideas. Organizations do 'not need to reinvent what others have done: Today's rallying cry is "acquire, adapt, and advance." Benchmarking, Kaizen, partnership and alliance, etc., are used in this regard.

Core competencies

Core competencies describe what the organization is specially or uniquely capable of doing. Innovation, quality, and cost leadership are the three most important factors to achieve competitive advantage and business excellence. All these depend on the quality of an organization's human resource and their competitive intelligence (Porter 1985).

Human capital

Human capital is the combined knowledge, skill, innovativeness, and ability of the organization's individuals to meet the task at hand, including values, culture, and philosophy. This includes knowledge, wisdom, expertise, intuition, and the ability of individuals to carry out value creating tasks and goals. Human capital is the property of individuals. Also, human capital includes human resources with the organization and also customers and suppliers of the organization (Stewart 1997).

Intellectual capital

IC (intellectual capital) is "the intangible assets – knowledge, information, experiences, intellectual property – that can be put to use to create wealth".

Information technology

IT (information technology) is redefining the way ideas grow and business organizations work. Three fundamental changes are foreseen in the wake of this technology revolution.
- Information is going to become knowledge.
- The digital will replace the physical.
- Everyone and everything will get connected, eventually.

107

Digital library and e-learning

Business organizations to sustain and grow in this knowledge-based, dynamic and competitive environment have started establishing digital library, network-based information and knowledge management systems for effective dissemination and sharing of information, and knowledge among their human resource for developing the human capital (intellectual assets). Also, one of the most striking workplace phenomena in the 21st century is the unrelenting demand for increased knowledge and speed of learning. Corporate e-learning in the United States alone is $ 15 billion in 2005. Organizations now invest in a wide variety of rapidly diversifying technology-based learning methods. They are generally customized, digitized, and optimized for each individual, thanks to the ever increasing power of technology and its growing applicability to learning in the organization

Knowledge management

There is a growing realization about KM (knowledge management) as the enabler of innovation and learning as well as national GDP (gross domestic product) (Boisot 1998).

KM is defined as "management of organizational knowledge for creating business value and generating competitive advantage". Knowledge is the outcome of learning. It consists of processes that facilitate generating, processing, sharing, using, and storing knowledge. It is an organization-wide effort for capturing and synthesizing information from which knowledge can be created, stored, and shared.

- *Explicit knowledge* – Knowledge represented in documents, books, e-mail and databases.
- *Embedded knowledge* – Organizational knowledge found in business processes, products and services.
- *Tacit knowledge* – Undocumented knowledge that is captured during business processes by knowledge workers.

L&T– knowledge driven organization

L&T (Larsen & Toubro Ltd) is a technology-driven engineering and construction organization, and one of the largest companies in 'India's private sector. A strong, customer-focused approach and the constant quest for top-class quality have enabled the Company to attain and sustain leadership in its major lines of business across seven decades. L&T believes that progress must necessarily be achieved in harmony with the environment.. The organization culture, vision, strategy, structure and the people play the pivotal role in building a learning organization in L&T. *People are the prime movers.* A commitment to continuous learning constitute an integral part of the corporate *vision.*

L&T shall be a professionally-managed
Indian multinational,
committed to total customer satisfaction
and enhancing shareholder value.

L&T-ites shall be an innovative,
entrepreneurial and empowered team
constantly creating value
and attaining global benchmarks.

L&T shall foster a culture of caring,
trust and continous learning
while meeting expectations of
employees, stakeholders and society.

LARSEN & TOUBRO LIMITED

108

L&T won 'Learning Culture' award 2005

(Mr. Y.M. Deosthalee (left) L&T CFO & Board of Director receiving the Learning Culture Award from Mr. Dilip Valse Patil. Dr. Peter Senge is at right.)

Mumbai, 25 November 2005 L&T won professional recognition for its unique culture and ethos by securing The Economic Times-Indira Group of Institutes Award for 'Organizations that Create a Learning Culture'. Mr Y M Deosthalee, Chief Financial Officer and Member of the Board, received the award from Dr Peter Senge (founding Chairperson of the Society for Organizational Learning, noted author and business strategist) and Mr Dilip Valse Patil (Minister of State for Education, Energy and Power), at a function in Mumbai on 21 November 2005. *The award citation noted that L&T's approach to organizational development has been benchmarked by the industry, and would help create an organization that encourages personal mastery in competence building.* The award presentation was part of a 'Strategy Summit' organized by the Indira Group.

LDC (Library and Documentation Centre) is one among the vital departments of L&T ECC Division. Its role is to enable development of employees through organizational learning - value added information services and knowledge sharing efforts. Its prime responsibilities explicitly include "conceptualize, plan and develop a world-class library and documentation centre – a corporate learning centre for information and knowledge management in the field of construction industry with all types of information resources and facilities to access and disseminate information and knowledge using information technology, to enhance the individual competencies and there by the core competencies of the organization. Develop a data bank of all techno-economic information for learning quickly from past experience and excel. Contribute for building a learning orgnization and living company".

Networking
ECC is members of about 97 national and international professional / trade organizations, institutions, societies and all these memberships are processed through LDC. As a result of this an effective networking is prevailing between these organizations and the company. About 90% of these communication and information services are being received through e-format and disseminated widely through network to the concerned users on a continuous basis.

Organizational learning initiatives

Good to great - learning to win by sharing
A world-class digital library, information and knowledge management initiative is a primary enabler in transforming an organization from good to great. Helps for learning quickly from past experiences and excel. Also, contribute for building a knowledge-based learning organization and living company. These organizational learning initiatives are primarily directed towards enhancing the individual competencies

109

(skills, knowledge and attitudes) of the employees and there by the core competencies, competitive intelligence, EVA (economic value addition), total customer satisfaction, shareholders value and maximization of resources, to sustain and grow in this dynamic environment.

Digital library

The head quarters, all the 14 SBUs, departments, offices, 7 regional offices and about 375 projects job sites, user community of about 9500 comprising 75% engineers, spread across the country and abroad are well connected by EIP (enterprise information portal) - URLs: http://www.lntecc.com and http://km.lntecc.com and Intranet URL: http://172.31.19.101 for information and knowledge sharing — success stories, best practices, experiences, etc. This promotes a culture and faster transfer of best proven practices and insights and helps to learn quickly from past experience.
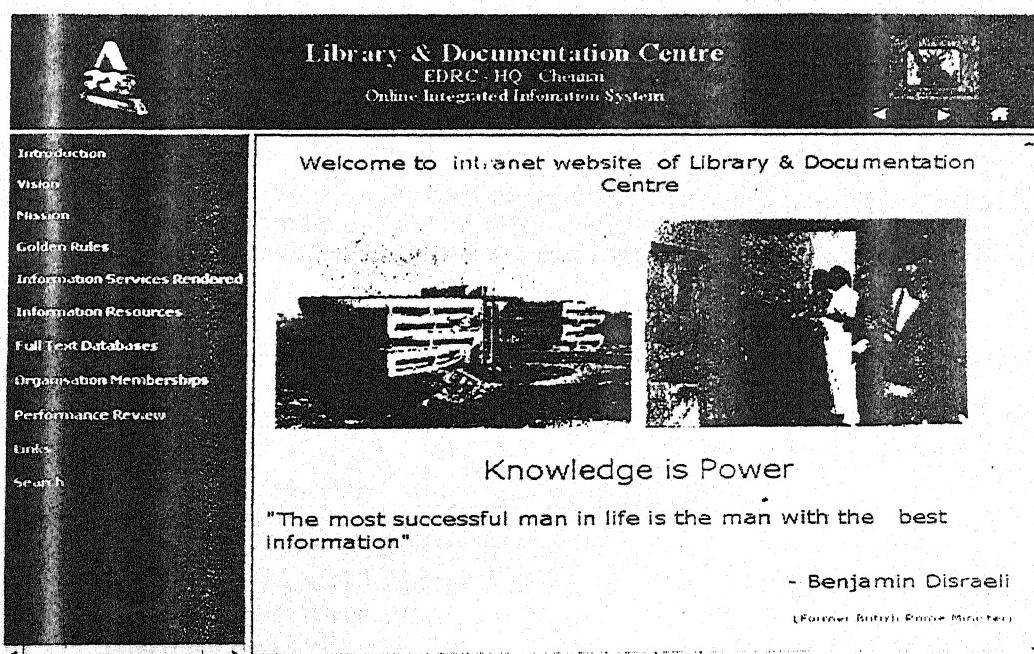
Figure 1 L&T ECCD digital library web page

Digital library resources
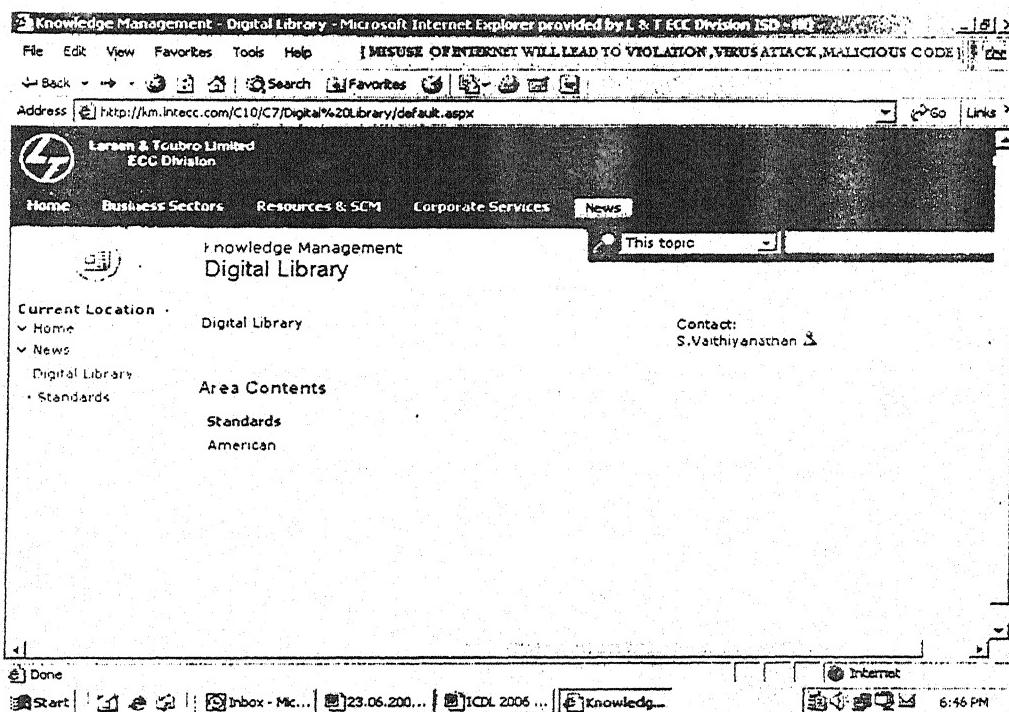
Figure 2 L&T ECCD enterprise information portal page

E-collections form the basic building blocks for the digital library. They are the digitized documents that are available for on-line access to the users. They are indexed, catalogued, and categorized. They present an infinite yet selectable array of information in a format that is easily navigated and harnessed for competitive advantage.

E-eatabases
E-books
E-standards
E-journals
E-thesis / project works
E-reports
E-articles (technical / management)
E-project profiles and design documents
E-business information service (sales leads)
E-learning & e-training (CD-ROMs, CBTs, video and audio tapes and slides)

E-in-house publications

Articles that are published in in-house publication like *ECC Concord* – the technical magazine since 1978 are made available in e-format for on-line full-text access by title, author/s, keywords, subject wise.
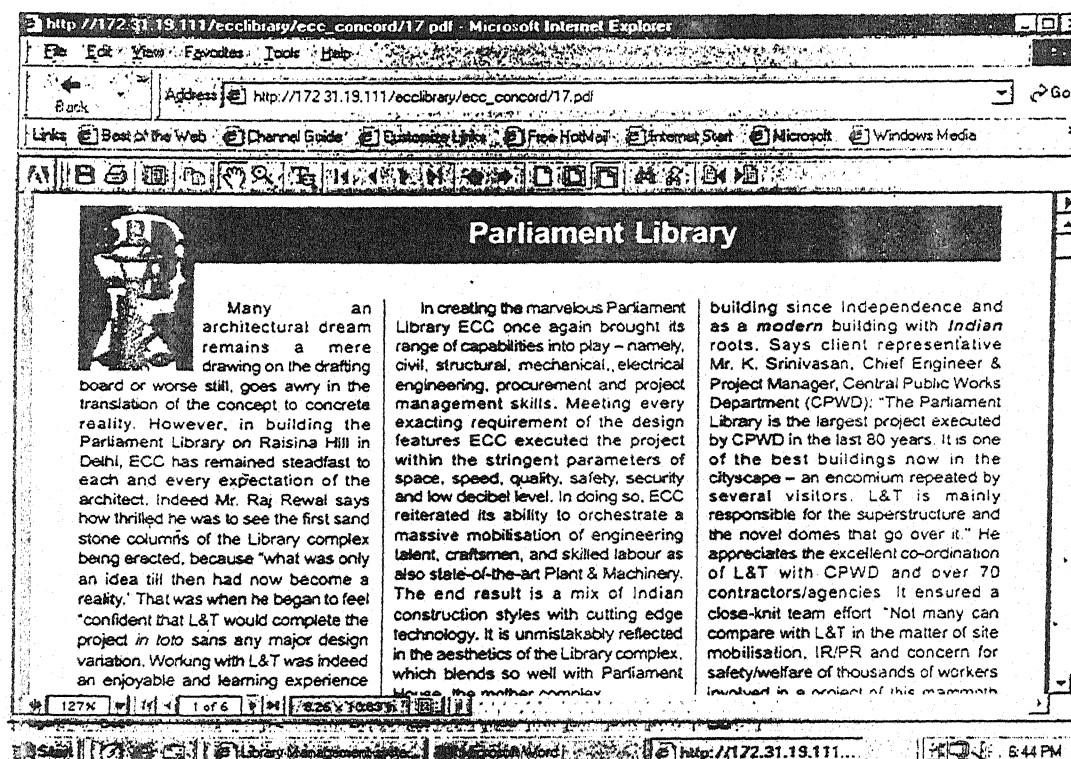
Figure 3 ECC Concord – in-house publication access page

Project profiles: transforming information and knowledge into intelligence

The EDRC projects design data documents and profiles of the completed projects serves as a useful reference material for pre-qualification purposes, when the company bid for new projects. The profiles contain success stories, best practices, experiences, and the data pertaining to new technologies adopted, improvements made in the existing methods, cost control measures adopted, etc., which are valuable tools during the execution of similar projects in future. They are created in standard e-format with complete details captured under 16 headings and are available for easy access and retrieval via intranet.
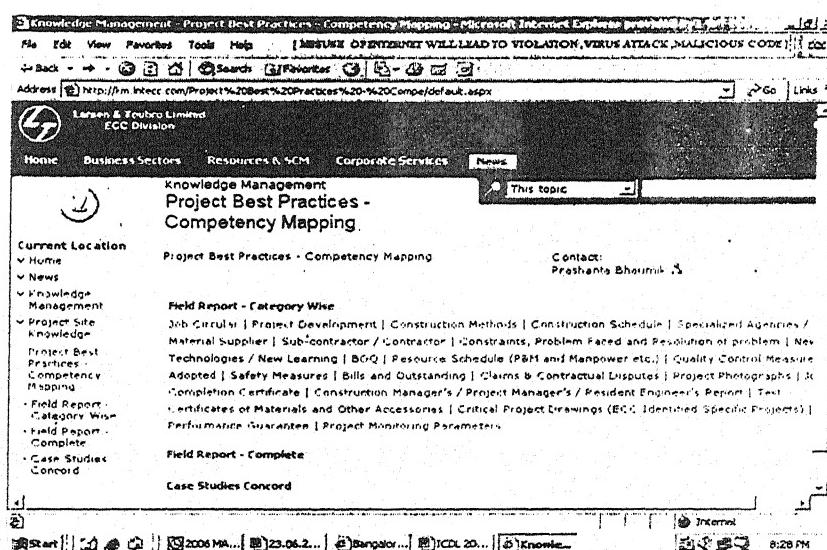
Figure 4 L&T ECCD EIP projects profiles access page
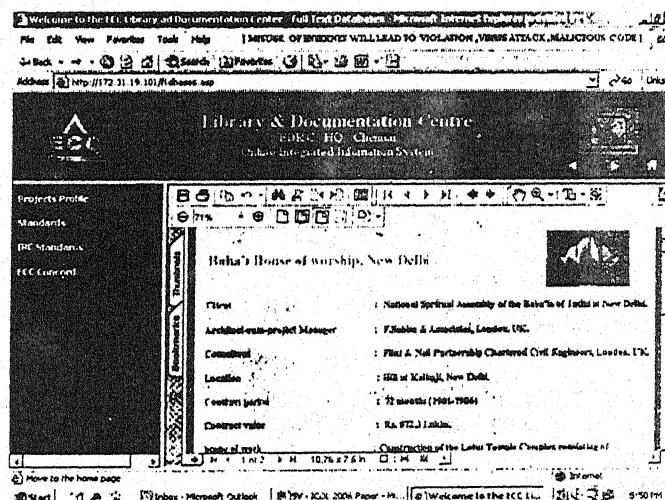
Figure 5 L&T ECCD project profiles access full-text access page

One-engineer – one journal review

LDC in association with SBUs and department heads introduced the one-engineer – one journal r
system. Under this a team of engineers have been identified. An engineer adopts a journal and reviev
issues depending on the periodicity. Selects articles of interest to the organization keeping the str
planning and the technology requirements. The bibliographic details along with abstracts and full-te
select articles are made available for on-line access by subject wise by all the employees – individual lea
transformed into team learning and organizational wide learning.

113

Soil investigation reports

Full-texts of 541 soil investigation reports of various projects executed in the country and abroad are made available in the e-format and organized country wise. They serve as a valuable source to know the soil conditions of the particular place while taking up the new projects.
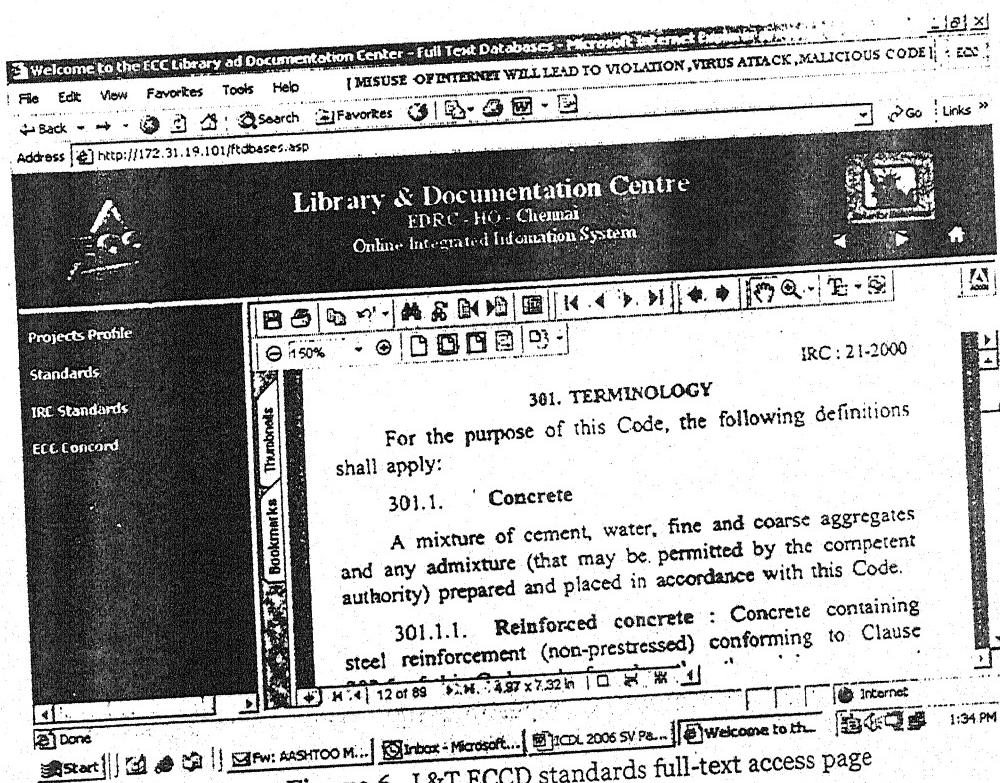
Figure 6   L&T ECCD standards full-text access page

New projects information sales leads: market growth enabler

LDC is compiling and circulating information on new projects in the offing, government policies and moves, economic, political and socio trends from information resources like news papers, periodicals and internet and are digitized and disseminated though e-mail to all the Department heads and managers cadres about 900 people. This is a daily service. This value-added information service helps the top management to know the market potential in each core sector and to take strategic policy decisions accordingly in targeting the potential areas of business in tune with the changing environment. These information stored in the EIP (enterprise information portal) by core sectors-wise and subject-wise for on-line access by all the employees.

114

Figure 7   L&T ECCD sales leads

Figure 8 L&T ECCD books search page

E-Learning and e-training: empowers employees with knowledge sharing

LDC has separate facility for self-learning packages and are extensively used. A good collection of self-learning packages CD-ROMs, CBITs, computer softwares, video tapes and audio tapes on a wide a range of subjects – leadership, finance, team building, communication, time management, stress management, project management, empowerment, ISO 9000, TQM, civil, electrical and mechanical engineering are available. These e-information resources have distinct learning advantages by the users leading to self-development and benchmarking.
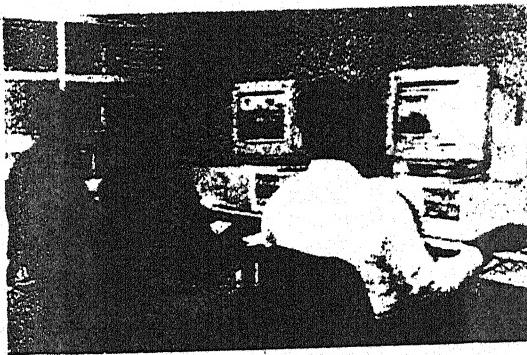
Figure 9 L&T ECCD e-learning / training access page

116
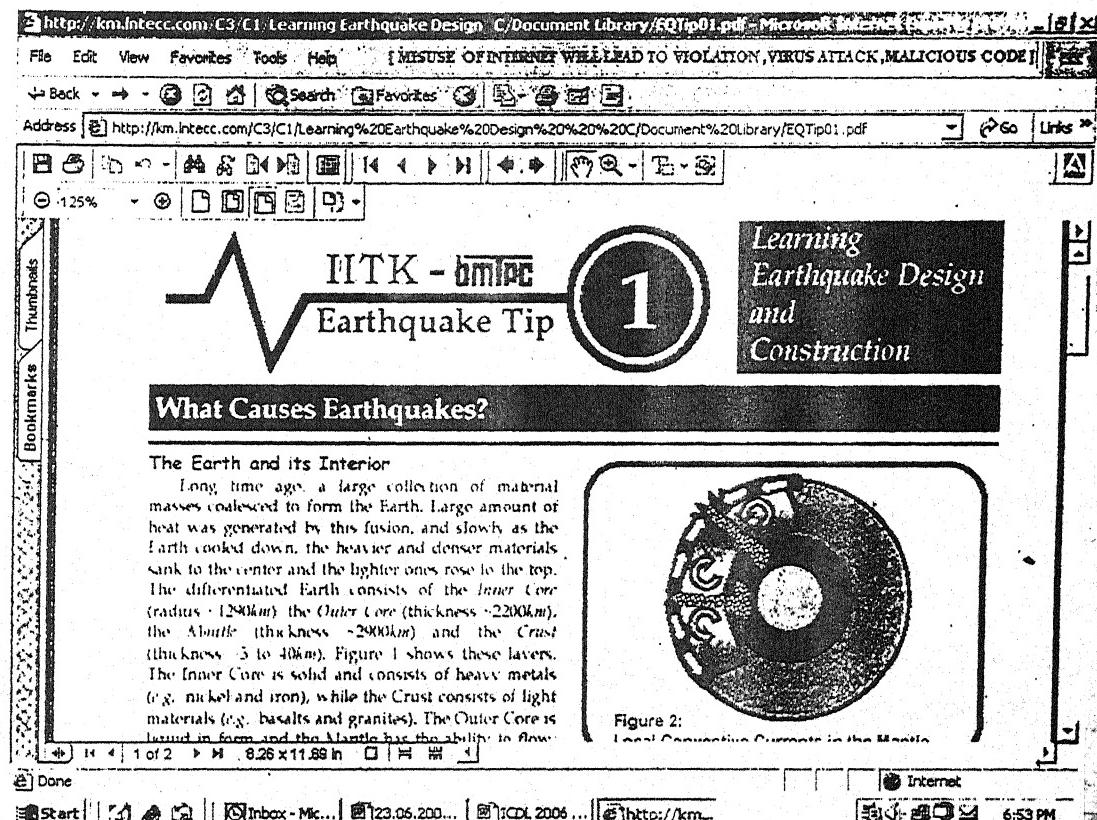
Figure 10 E-Reports full text access page
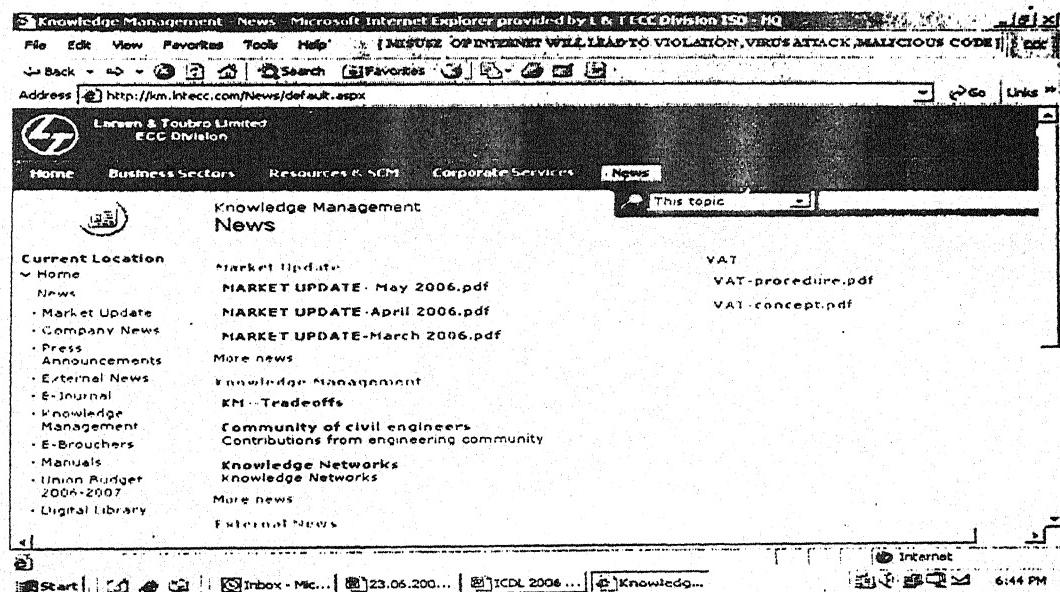
Figure 11 L&T ECCD E-resources access page

Conclusion

'The fundamental cure for poverty is not money but knowledge'.

— Sir William Arthur Lewis

Information and knowledge is the key asset of an organization, which gives sustainable competitive advantage. The organizational learning brings more transparency in organization culture and operations. This results in continuous innovation in services. In today's highly competitive business environment, learning organizations hold a significant competitive advantage. Their ability to harness the power of learning at all levels — individual, team, and organizational — enables them to rapidly leverage new knowledge into new products and services, new marketing strategies, and new ways of leading the learning revolution. Learning organizations needs the commitment, modelling, and involvement of all top leaders.

If knowledge is power, then the release of knowledge is the release of power. Whether from person to person or from computer to person, such a release is an infusion of energy and strength into an organization. This is what learning organizations thrive on. In a learning organization, the journey never ends. Initiatives such as continuous process improvement (Kaizen), Benchmarking are just part of a learning organization's overall goal to increase productivity, efficiency, and effectiveness. Things move faster every day. 'Like it or not, the years ahead will be an era of change and competition'. 'The only thing one can predict about the future is change. To keep up, we must stay on our toes and adapt, change, and learn at every opportunity, everyday, All the time'.

Major Websites

URL

www.start4all.com	- Newspapers
www.unweb.carl.org	- Current Contact
www.isidev.delhi.nic.in	- Institute for Studies in Industrial Dev.
www.sosig.ac.uk	- Social Science Information Gateway
www.asiatica.org/publications/jsaws	– Asiatic Society
www.unesco.org	- Unesco
www.loc.gov	- Library of Congress
www.worldbank.org	- World Bank
www.gale.com	-Gale Directory
www.Lse.ac.uk	- London Sch. of Eco.
www.ids.ac.uk	- Devline. Instt. of Dev. St.(UK)
www.unsystem.org	- United Nations
www.columbia.edu/cu/libraries/indiv/area/sarai	-South Asia Resources
www.ilo.org	-ILO
www.hwilson.com	-H.W.Wilson co. Journals.
www.censusindia.net	- census
www.igidr.ac.in	- Indira Gandhi Institute for Dev. Research
www.bbc.co.uk.	- BBC

119

The Internet for Social Scientists

by Craig McKie and Paul de Guchteneire

What is the Internet?

The Internet is simply a very large number of interconnected computers. These computers, (each with its own address or 'IP number'), are connected in a single global co-operative by wire, by fibre optic cables, by satellite links or by telephone lines. The manner of connection is not nearly so important as the fact that each member computer can exchange large amounts of digitised information freely across the world with any or all other connected computers. Distance has thus ceased to be the decisive factor in human intellectual life. As a result, the social scientist can ill afford to ignore the power of the Internet to both disseminate and to collect and sift information in all its forms.

From the point of view of the social scientist, the Internet offers access to unimaginably large amounts of information, data, and interpretative material in a timely, cost-effective, and comfortable manner. Further, the user can easily become an active contributor to the body of knowledge on offer in the world with very little additional effort. Contributions can be made through the exchange of electronic mail with distant colleagues, through postings to the USENET newsgroups, and to topic-particular e-mail discussion groups (listservs). Integrated with everyday work, the Internet offers an incomparable opportunity to actively participate in the accumulation and dissemination of a truly global body of professional social science knowledge, expertise, and opinion.

Using the Internet is getting easier all the time, in no small measure because of the rapid adoption of the World Wide Web as a presentation standard for all manner of digital materials including text, graphics, programmes, and even audio and video across the Internet. Use of 'the Web' is based on a 'point and click ' approach in its full graphics form and the use of the keyboard arrow keys in the less attractive text-only mode. Both methods are easy to master, hard to break, and are forgiving in the extreme for novices. And, when a resource is located, its location can be saved so that the address (or Uniform Resource

Locator, the URL) need never be written down or committed to memory. The saved URL, which is stored together with a descriptive title of the user's choosing, can be easy recalled and used. The Web itself has become the contemporary equivalent of the great library at Alexandria, though it is much more robust than its predecessor because it is not flammable, nor is it subject to the intrusive influence of any single government. The Web is a liberated zone of free and, if necessary, anonymous expression.

How can social scientists benefit from the Internet....

Social Scientists can benefit in many ways from the use of the Internet toolset. In general, the Internet delivers better, faster, more timely communication with colleagues and sources of information than has ever before been available to social scientists. It offers enhanced collaboration, better dissemination of one's information, instantaneous peer review, and low barriers to publication of drafts and requests for comments. It can also be used to acquire huge public datasets from a multitude of national data services and international agencies. Fast searches of recent periodical literature are also available free of charge as are current information sources on contemporary geopolitical events. Taken together, these tools allow social scientists to become active participants in the events which shape and illuminate contemporary social processes and discourse. They are a marvellous addition to the set of skills offered to the world by social science as an institution.

In the future, new tools will be added to this toolkit. Already, experiments with slow scan television carried over the Internet have made it possible for professors in one continent to conduct interviews of colleagues in another 'live' in the classrooms of each. Though the sound quality is excellent, the slow scan television is yet not up to commercial standards but the spontaneity of such exchanges of views renders the video quality of little significance. Also available are software packages with enable two-way telephone-like voice conversations between colleagues over the Internet which are free of charge and scrambled to prevent interception. In general, these new and experimental software advances are free of charge (for example, CuSEEme television, PGPphone, and Nautilus)

The MOST Clearing House on the Internet.

MOST makes use of the full potential of the Internet to disseminate information from the programme and to facilitate the co-operation of researchers in the joined international projects. Since the distance between the partners in MOST projects is usually big, some projects are even world-wide, co-operation through the Internet has become the only reasonable alternative for efficient communication and collaboration. The MOST Clearing House provides several tools for collaboration such as e-mail discussion lists, the circulation of draft documents and the facilities to give feedback to working papers. For the broader scientific public the Clearinghouse provides access to all publications and documents of the programme in the available languages, and to an Events Calendar that features up-to-date announcements on upcoming events. A publicly accessible databank on Best Practices in policy-making is in preparation. Finally, several discussion lists in which interested scholars can discuss the themes and projects of MOST are being established.

The MOST Clearinghouse can be reached at http://www.unesco.org/most

Tools of the Internet....

Electronic Mail - e-mail

allows you to exchange information, files, manuscripts rapidly and effectively (including files from the common word processing packages if suitably encoded). Many exchanges of

121

messages can take place during a working day. The effect is that of a conversation between interested co-workers. Also possible is active collaboration on a single project between several colleagues thousands of kilometres apart who may never have actually met each other. New collaborations become feasible; old ones become closer and more continuous. Because time zone differences can be overcome ('conversations' need not be confined to times when both parties are 'logged on'), the range of possible collaborators is greatly expanded. Every user has his or her own unique e-mail address. It looks like *someone@somewhere.com*. The first term is the user's logon name and the second is his or her home machine's name. Use is as simple as composing a memo and sending it off to its destination with a few key strokes. It is however mandatory that you know the correct e-mail address of your intended recipient.

· Listservs

are dissemination engines. They are e-mailing lists which can send single e-mail messages (including many sorts of files if desired) to an unlimited number of recipients who are members of a list. Any member of a list can, with one message, address all of the members of the list. Listservs are maintained at a single location and to write to the entire list, all that is required is a single message to the listserv. Listservs can be started on almost any Internet host computer that has the required (free) software installed. In principle, any group of e-mailers who share a common interest can start a listserv on any subject imaginable. Lists can be open to all or closed to all but welcome applicants. Postings can be moderated (reviewed by an editor) or unmoderated; additionally, e-mail postings to the listserv can be encrypted so that the source of a posting can be truly anonymised if desired. In addition, there are features which allow compilation of daily message packages rather than or in addition to allowing the individual retransmission of submissions as soon as they arrive. Some list owners also maintain archives of submissions.

· Newsgroups

Another way in which information is circulated by Internet users is by posting to and reading of the USENET newsgroups (of which there are now thousands). Newsgroups are topic specific. One simply posts information or a question in the form of an e-mail message to a newsgroup or newsgroups which seem appropriate for your interests. Readers 'subscribe' to newsgroups they wish to read using reader software such as *tin*. Since posting and reading are independent actions, and there is no obligation to do either, it is never certain who may read your posting. In practice, some newsgroups are very well read, particularly where they have come to be a place for debates on topics that attract a large and active audience. You may post questions and receive advice from others very quickly, sometimes in a matter of minutes. In essence, strangers (who may post anonymously should they wish) may become your information agents in many remote locations simultaneously.

· The World Wide Web

is the most sophisticated venue of information exchange on the Internet. The user has access to millions of 'Web pages' which may contain data sources, information and/or other onward 'links'. The resources of Web pages maybe viewed with a Web browser (such as *Netscape*) or textually (with a text based browser such as *Lynx*). It helps to know where to start. One good place to begin is *Yahoo* in California, a site started by two students at Stanford University. Its address is *http://www.yahoo.com*. It has a well developed social science component of useful links. All addresses on the web are in http format (hypertext transport protocol).

· Telnet

is the modern traveller's friend. When you have an Internet account and you are travelling,

messages can take place during a working day. The effect is that of a conversation between interested co-workers. Also possible is active collaboration on a single project between several colleagues thousands of kilometres apart who may never have actually met each other. New collaborations become feasible; old ones become closer and more continuous. Because time zone differences can be overcome ('conversations' need not be confined to times when both parties are 'logged on'), the range of possible collaborators is greatly expanded. Every user has his or her own unique e-mail address. It looks like *someone@somewhere.com*. The first term is the user's logon name and the second is his or her home machine's name. Use is as simple as composing a memo and sending it off to its destination with a few key strokes. It is however mandatory that you know the correct e-mail address of your intended recipient.

· Listservs

are dissemination engines. They are e-mailing lists which can send single e-mail messages (including many sorts of files if desired) to an unlimited number of recipients who are members of a list. Any member of a list can, with one message, address all of the members of the list. Listservs are maintained at a single location and to write to the entire list, all that is required is a single message to the listserv. Listservs can be started on almost any Internet host computer that has the required (free) software installed. In principle, any group of e-mailers who share a common interest can start a listserv on any subject imaginable. Lists can be open to all or closed to all but welcome applicants. Postings can be moderated (reviewed by an editor) or unmoderated; additionally, e-mail postings to the listserv can be encrypted so that the source of a posting can be truly anonymised if desired. In addition, there are features which allow compilation of daily message packages rather than or in addition to allowing the individual retransmission of submissions as soon as they arrive. Some list owners also maintain archives of submissions.

· Newsgroups

Another way in which information is circulated by Internet users is by posting to and reading of the USENET newsgroups (of which there are now thousands). Newsgroups are topic specific. One simply posts information or a question in the form of an e-mail message to a newsgroup or newsgroups which seem appropriate for your interests. Readers 'subscribe' to newsgroups they wish to read using reader software such as *tin*. Since posting and reading are independent actions, and there is no obligation to do either, it is never certain who may read your posting. In practice, some newsgroups are very well read, particularly where they have come to be a place for debates on topics that attract a large and active audience. You may post questions and receive advice from others very quickly, sometimes in a matter of minutes. In essence, strangers (who may post anonymously should they wish) may become your information agents in many remote locations simultaneously.

· The World Wide Web

is the most sophisticated venue of information exchange on the Internet. The user has access to millions of 'Web pages' which may contain data sources, information and/or other onward 'links'. The resources of Web pages maybe viewed with a Web browser (such as *Netscape*) or textually (with a text based browser such as *Lynx*). It helps to know where to start. One good place to begin is *Yahoo* in California, a site started by two students at Stanford University. Its address is *http://www.yahoo.com*. It has a well developed social science component of useful links. All addresses on the web are in http format (hypertext transport protocol).

· Telnet

is the modern traveller's friend. When you have an Internet account and you are travelling,

a colleague with access to an Internet account wherever in the world you are may help you to *telnet* to your home Internet account. You may then logon and use it as if you were in your own office, reading your e-mail for instance. There are no additional charges for this service no matter how distant the telnet connection is nor how long it lasts; its part of the toolkit.

· The older tools

There are several older Internet tools. Without going into detail here, they function to distribute free software, information and allow simultaneous real time conversations between dispersed participants. These tools are still valuable but their functions tend now to be 'buried' in other more sophisticated software packages and are thus hidden from view.

Where to Start....

· Getting Connected

Every user needs access to the Internet by means of a 'member' computer. Many professionals choose to do this by means of a personal computer, a modem and a telephone connection to an Internet 'host' computer. In the academic context, a connection may be provided in a number of different ways: Ethernet cable, internal telephone systems, fibre optic connections, and so on. Each, though, provides a 'pipe' between the Internet and the user's machine. Connections come in many gradations of quality. Since faster is better, faster also tends to involve more expense. In the future, very fast connections (by which is meant high volume connections) will become commonplace, perhaps through the shared use of cable that now is only used for the delivery of television cable services to households. Obviously, the quality and quantity of high technology infrastructure available to the user is a function of the national wealth of his or her country.

· Learning the skills

One of the charms of the World Wide Web is that it is to a large and growing extent self-documenting. Once 'there', you can easily locate free learning materials to further develop your skills. However, getting 'there' is still a formidable hurdle. While it is possible to buy printed reference material, it is best to learn from an experienced user. Becoming an 'information apprentice' seems, on experience, the best way to learn the tricks, the shortcuts and the 'magic' instructions. In part this is so because the Internet runs for the most part using the UNIX operating system at its base. This system is as arcane as it is beloved of Internet programmers; commands are not intuitive and often could never be guessed by the novice. Direct instruction by a trusted and knowledgeable colleague is the best solution. Often, it takes just a few hours of tutoring to overcome the initial barriers and enter the self-documenting zone of the Internet. While this initial step requires a degree of faith, the novice will be richly rewarded by newly granted access to the developing world of shared data, knowledge, research advice, publication and hopefully wisdom.

... and Onwards...

· Routine use

Once you have reached your goal of easy use, you may well find that the Internet toolset becomes a part of everyday life, moving far past professional use into all aspects of personal existence. Having easy access to the world's newswire services in real time and major newspapers and newsmagazines before they are published tends to do that. Embracing this sort of change in learning styles is deeply enriching on a personal level and brings us all a

123

little closer together.

· Your Own Mailing list to establish and maintain contact

Once you have identified members of your own particular Internet community, no matter where they might happen to live in the world, it is possible for you to establish your own mailing list or *listserv* to remail messages from any member of your list to all members of the list. USENET postings and listserv messages inviting colleagues to join a new listserv on some narrow specialised topic is a normal part of the world's information traffic nowadays.

...to Augmented Services.

· Setting up your own Internet Services

You may wish at some time to actually establish and maintain your Internet services on a machine over which you have exclusive control. In order to do this, many elements of hardware, software, and registration have to be put in place. For instance, you must obtain a computer over which you have authority and which has the technical capacity to serve as an Internet host. It need not necessarily be powerful and expensive though each helps in some respects. Such a computer must be provided with a high speed network feed from some other member machine on the Internet. It must also be given a unique IP address by some registration authority such as the INTERNIC. In addition, it must have access electronically to a router computer which has a current copy of the IP addresses of other Internet member computers. Further, specialised software such as UNIX, Web server software, and e-mail software must be installed and properly configured. In general, a high level of sophisticated knowledge is required to carry out these tasks. Should you choose to assume this responsibility, you will also have to take necessary precautions to guard the security of your site since attacks by hackers can be confidently predicted. While there are security software tools and expertise available to you, you may well find that the routine keeping of system logs is a requirement and that trouble can develop at any time of day or night. (A low budget version of this software universe called LINUX is available for modest personal computers running in a low volume context.) You, as system administrator must also decide how external users will access your service, whether by *telnet*, dial up access, or exclusively as external Internet Web users visiting from other Internet hosts. Committing to offer dial-up services to others may itself pose formidable technical and expense challenges.

· Maintaining a WWW site to publish your own information, reaching a global audience

Once you have access to an Internet service, whether run by yourself or by others, and once Web server software is installed, it is quite simple to begin to publicise information on any topic through your own Web page. Editing tools are available on the Web to help you master and construct HTML documents for your web site. One of the best ways of learning how to do this is to look at the HTML code which other users are employing on their web pages. All Web browsers allow you to view the source coding for others' pages. Copying of features which appeal to you is an easy and effective way to get started. It is also useful to register your web pages once they are constructed and available with some of the many registries of Web pages. This will allow users around the world to locate and visit your site based on your brief description of its contents. In addition, there are page-seeking Web robots (such as *Alta Vista*) which actively search out and catalogue the contents of new web pages all the time.

Other Resources you may offer

You may use your Web site to offer virtually any form of digital resource to other users. You

may for instance make available databases, papers you or others have published, software you wish to give away or announcements you may wish to publicise. The situation with respect to copyright on Web documents is not at all clear and there is little guidance available. In general, normal good citizenship on the Web includes not appropriating the work of others for your own gain without permission. Having said that however, it is the very essence of the Web that copying of files takes place without hindrance. Indeed, that is the main attraction of the World Wide Web as it has evolved.

Craig McKie
Carleton University, the Centre
for Online Studies International,
and Ingenia Communications Corporation,
Ottawa Canada
Email: cmckie@ccs.carleton.ca

Paul de Guchteneire UNESCO MOST
Fax + 33 1 45 68 57 24
E-mail p.deguchteneire@unesco.org

To MOST Clearing House Homepage

125

Management of digital resources

Kamal.K.Chauraasia

Introduction

Libraries are the storehouses of information and knowledge. Information is stored in many forms: print, audio and video, motion picture, slide microfilm/microfiche, etc. The sheer volume of paper information generated has perforce brought the shift to electronic media. In today's world of science and technology, the need of information is extreme and to satisfy the needs of this information, society is quite puzzled and fulfillment of these information needs requires use of computer and network technology and automating one's jobs, with the high speed internet connectivity.

Today library and information professionals are normally stressed with user demands regarding availability, storage and access of e resources. Use of information technologies has marked a tremendous impact over the all functions and services catered by the traditional library and information professionals. E-resources are now emerging as a vital source of information for all recent and nascent thoughts and ideas coming into existence in whatever area of research. Whole books and periodicals have started appearing (in digitized forms) on CD ROM discs and others. Emergence of Internet and www has provided a platform to display these resources globally.

There is vast amount of information, which needs to be accumulated, processed, organized and presented in the most usable form and is amenable to quick and user friendly retrieval. This is where the idea of electronic publishing emerge, thus e – publishing can be defined as "the publication process in which the manuscript are submitted in electronic format, edited and printed and even distributed to the users in the most usable form such as CD-ROMS, Diskettes or Internet or even local area networks".

Electronic products may include text, graphics, audio, video, numeric and textual databases. It can also be said that electronic publishing is dissemination of information in the electronic form. Some products of electronic publishing are as follows:

⇒ Electronic Journals
⇒ Electronic Books
⇒ Electronic Databases

Management

Management literary means here to manage tactfully. Effective management is a key to organizational success. But mis-management squanders our resources, creates chaos and disorder and endangers our well-being. Few definitions of management are mentioned here for better clarity of management:

Definitions:
Management is simply the process of decision-making and control over the action of human beings for the expressed purpose of attaining pre-determined goals. - *Stanley Vance*

* Librarian, Moti Lal Rastogi School of Managementt. Lucknow-8. Email: klib_522@yahoo.com

126

Management is the process by which managers create, direct, maintain and operate purposive organization through systematic, coordinated, cooperative human effort. *-Dalton E McFarland*

Importance of management in libraries

Books alone do not make a library. Similarly a good collection of books also is not enough to ensure the successful working of a library. The success of library depends in large measure upon the persons who are responsible for their use and distribution and fulfilling its objectives. The collection of a library must have its effective use in the hands of readers. The person who takes responsibility for this transformation is the librarian; though more often than not he is decorated with some such prefix as 'chief' or 'city' or 'university'.

The best-stocked library cannot give anything like hundred percent services to readers unless it does possess a keen, efficient and trained staff that should be able to exploit the stock to its fullest advantage. The major benefits of management can be summarized in following manner.

- Management tries to make effective utilization of various resources in the library.
- Management is helpful in development of various resources in the library.
- Management is helpful in applying innovations in routine activities as today changes are occurring at a fast rate and they need to be incorporated to keep the organization alive.
- Management secures stability in society by changing and modifying the resources in accordance with the changing environment of the society.

Information technology- a strategic resource

IT is becoming a strategic resource. The capability of IT is continually increasing and the political, social and economical context of its use is undergoing simultaneous radical changes. This combination of contextual forces for change and technological advance has created considerable rhetoric.

IT can be used strategically in at least four different ways:

- To gain competitive advantage
- To improve productivity and performance.
- To facilitate new ways of managing and organizing.
- To develop new dimensions in concerned field or area.

Of course the boundaries between these uses are imprecise and the categories overlap, but they each represent different intents. However, it could be claimed that computing, data processing, information services etc., have always been pursuing such goals. The difference is that the multiplication and convergence of information technology provide more potential and the structural changes in economies, industries and organizations more opportunities. Thus, a new breed of information systems-strategic information system has arrived.

Managing E-Resources

Libraries generally did not require any specific arrangement for print resources for offering, receiving and processing since these processes were well established. But the processes for new medium are still under the stage of experimentation especially in libraries in our country. The libraries are still trying to grapple with the issues of

127

copyright, licensing system compatibilities, access speed and time, downloading, book marking etc. Each of these issues are complex and at the same time interdependent. To avoid ambiguities, it would be necessary to have a proper planning and management mechanisms. Libraries need to have kept in mind the following aspects regarding the e-resources:

1) **Multiple vendors offering same titles** - libraries must choose the best provider.
2) **Vendor Viability** – reliability & reputation of the vendor.
3) **Licensing** - licensing determines what may be accessed by whom, and for what length; licenses also give provisions for what happens when the library no longer wishes to subscribe (access to material "paid" for during the subscription time period); a license must be signed for each availability option except the freely-available online resources and the pay-per-view; license negotiation is time-consuming and must be done carefully as it is a legally-binding contract.
4) **Costs** - while one would think electronic resources would be less costly because delivery is available in online format (no binding, checking-in of print versions, etc.), many publishers in fact charge as much if not more for electronic resources than their print counterparts.
5) **Varied Content Availability** - Titles can be added or dropped by a vendor at anytime. A library may subscribe to an e-journal, online database, or other mechanism at one point to gain access to titles not previously held. However, it is at the publisher's whims that the content is available at any given time. Some publishers put restrictions on availability, making content accessible only after an prohibition period.
6) **Multiple formats** - most e-journals offer the full-text of their materials in HTML format or PDF format. In some cases, though, format options may cause problems for access and readability. Multiple formats may also refer to access to both print and electronic versions. Libraries are now taking a serious look at having the same article in multiple formats -- is it considered an efficient use of money to have duplicates?
7) **Collection Development priorities** - libraries have had to change their perspectives on how they are developing their materials collections. E-Journals, E-Books, and electronic resources are now taking a bigger role in this process, particularly given patron expectations. This has an impact on budgeting and service.
8) **Accessibility** - one of the biggest management issues from the library-side is that of offering access to these resources. What is the appropriate mechanism for providing an entry point to e-resources? Solutions have been offered from various perspectives.
- **Vendors/ aggregators/ Free online resources/ publishers directly:**
  Online database subscription through commercial vendor, direct subscription from the publisher, freely available online resources and aggregators are the prominent medium to access the e-resources in a library. Each provides slightly different capabilities, but all a searchable interface for locating specific types of resources to see if they are available full-text online. Libraries subscribe to these services; may host the software on their own servers or contract with the provider for remote hosting; and configure these products to tailor them to the local library environment. The benefits are that the provider often has built-in helper applications to simplify the configuration process of loading journal titles and also providing detailed statistical reports.

- **In-House Software Creations**: Some libraries began providing access to e-resources prior before companies were developed to fulfill the need. These in-house products are more customized to the library, but also require more staff time to maintain. One example of in-house creations is Journal Finder http://journalfinder.uncg.edu
- **The OPAC**: Some libraries have looked to the OPAC to provide access to e-journals. In the vein of one-stop-shopping, bibliographic records for e-journal titles have been included in the online catalog, with patrons expected to search for an electronic journal in the same manner as they would for a title to which the library subscribed to in print. Concerns have been raised regarding the ability to keep the catalog current, the idea that e-journals are "different" than print materials and so their accessibility should be managed in other ways, and a host of other issues.

Here, I have tried to give a glimpse of the major managerial issues related to e-resources in a library. We have now evolved to a point at which so much R&D information are available in electronic form on the web, greatly extending the availability of these resources over time and distance. But it remains to be seen if the success of this new delivery means will ultimately fatally undermine the status of the traditional print format, as we have come to know it from generations. E-resources are now become essential part of library and information centers hence it is almost necessary for the librarian community to be well versed about the various issues related with its management and organization.

REFERENCES:

1. Franks, John. What is an Electronic Journal? Department of Math's, Northwestern University, Evanston, IL. January 1993.
2. Moshal, B.S. Management: theory and practice, New Delhi: Galgotia, 1998 pp 1-6
3. Munsi, usha mujoo, visualizing an Electronic Library/ Usha Mujoo-Munshi. IN Vistas in Library and Information Systems and Networks: a festschrift to professor S. M. Tripathi/ edited by Venugopal, M V [et. al]. -Agra: Y.K.Publshers, 1999
4. Prasad, L.M. Principles & Practice of Management. Sultan chand & sons: New Delhi, 1999 pp 7-15
5. PSG Kumar, Management of Library & Information centers (Paper V of UGC model curriculum), Delhi: B R Publishing House, 2003 pp 429-435
6. http://www.infolibrarian.com
7. http://www.library.uiuc.edu/icsu/hall.htm
8. http://www.ub.es/biblio/bid/06keefe2.htm

E-JOURNAL ACCESS: CHALLENGE BEFORE LIBRARIANS

PRAGYA PANDEY* & KAMAL KUMAR CHAURASIA**

INTRODUCTION

The term Electronic Journal refers to a journal which is available online and may or may not be associated with a traditional printed form. A simple assumption is that the transition from paper to electronic form will help in providing better services at a low cost. A range of innovations and features become possible with online journals those are either impossible or difficult to achieve with paper journals or even with electronic editions. In general, e-journals have few distinct features as they electronic journals can be searchable, and some are free too. Papers can include hypertext links to other resources. Few other journals offer links to resources beyond the journal.

Information professionals must have clear understanding of the information needs of the users and sound knowledge about information resources and how these resources are going to be used. This will result in building adequate resources to meet the mission and goals of the parent organization. Complex and dynamic nature of providing access to resources presents a challenge to a librarian. He must meet the challenge by developing a strong collection appropriate to the mission and goals of the parent body, serving the information needs of the users adequately.

E-journal access as seems for layman or the information seekers is not as easy as the traditional print journal use and access for the library and information professionals. Subscribing e-journal is not an easy task as it was in the case of print journal subscription. There are so many important issues relating to e-journal subscription and use in libraries. First thing involve in this process is to select the appropriate e-journals for both the user community and the institutional budget. Then acquisition of e-journals so as to provide the access to the users. Retrieving the information from e-journals also need some kind of expertise.

During the use and access of e-journals information users generally need special attention regarding the use pattern and ways to get the adequate information from e-journals. They should be well versed with the downloading techniques and laws relating to the intellectual property rights while accessing the e-journals and using its contents in their research outputs. The responsibility of library and information science professionals regarding the safe use of e-journals extended in this concern because the wrong use may harm both the subscribing library and the users too.

The following description will be helpful to the library and information science professionals about the effective use and access of e-journal, as the importance of these particular type of information source cant be ignored by the library and information science professionals, so as to accelerate the momentum of research and development in their concerned organizations.

1. SELECTION

Selection of electronic journals is the task in which a prompt analysis is required for the accurate selection of the necessary journals criteria can be summarized for the selection procedure:

⇒ Well-established multi-disciplinary resources with broad coverage were preferred over highly specialized sources targeted for specialists.

⇒ The electronic resources already on subscription in the beneficiary institutions were preferred over those, which are not being used, in any of the beneficiary institutions.

⇒ Resources that are 'electronic-only' were preferred over those that are print-based unless completely unavoidable.

⇒ Resources that are very important but highly cost-intensive were preferred over those,

which are less important or less-used but low cost.

⇒ Resources where electronic versions are made available free on subscription to their print versions were avoided as far as possible.

⇒ Selections were made on their usage / suitability to their respective institutions. Although above criterion is given but basically the points given below are to be considered foe the selection of the e-journals

Authority: - Find out evidence about the authors or producers of the sources of information to determine how far they are knowledgeable, reliable and truthful. Determine as to what is the subject expertise of the author, as given on the first page. In case, this information is not available on the first page, then one needs to trace backward in the URL (internet address) path to locate a page in the higher directory, giving the required information. Institutional affiliation can be identified in the e-mail address (.com for commercial products or commercially sponsored sites; .edu for educational or research organization; org for non-profit organization). If there were a personal name in URL (internet address), then it would represent personal home page without official sanction. Some of these pages are of high quality but it is rather difficult to determine their reliability. E-mail link used for sending queries and comments can bring more information that can be useful in evaluating it.

Scope: - Is the scope stated clearly, along with limitations and omissions, if any?
To what extent, the subject area is covered adequately at the required level of targeted customers?

Reliability: - Is it well compiled with information taken from valid sources of information? How accurate is the information?

Accuracy
1. One should look for evidence of the source being balanced.
2. For controversial issue, one should see if different opinions are presented fully? 3. Does it give bibliographical references to other authoritative sources?

Currency: - Currency and timeliness add to the value of resource. Dates in printed sources can be verified easily but for electronic resources, many items covered in web sites are based on the sources that are outdated editions. Therefore, in such cases, a user should seek the assistance of professional staff for finding out the currency of the sources as well as locating more current information. For Internet resources one should check:
What is the date of posting of information initially?
What is the date of revision, if applicable?
How current is the information given in the resource?
Is the information updated regularly?

Support And Backup: - For Electronic Journals, issue of support and backup is very important. So before selecting such a source one should find of what is the level of support and backup provided by the producer in the form of online help.

Functionality: - It has following issues to be concerned.
How is the website organized?
What is the style of writing is it appropriate to the target audience?
Is there a need for training of the end users and the staff?
Are there help files and functional links provided?

Format: - Does the format allow the documents to be viewed and downloaded in a variety of formats such as text or pdf?

131

Does it offer option of a database with or without text?
If it is also available in print, does the product offer original printed text along with the other illustrations?

Versions of the product: - Are the different versions of the product available?
If yes then do they contain the same information?

Technical Infrastructure: - Is the product compatible with the existing technical infrastructure of the library?
Does the library need to have specific software to access the database?

Terms of agreement under the license
What are the licensing requirement and liabilities?
What are the restrictions placed on access, copying and distribution?
How many users can access a product and download the required document simultaneously?

Archiving: - Does the vendor give right to the library to maintain access in perpetuity for the product to be archived?

Cost: - What is the cost of the product with different options as given Subscription with limited or unlimited access to the documents? Subscription for a limited number of users.

Prepaid searches: - Charges based on the amount of time accessed.
Free search with charges for viewing and downloading
What are the charges for the users at remote access?

Updating: - Is the product updated regularly? If yes what is the frequency of updating? Several other issues like quality of indexes generated and bibliographical control also, are to be considered while selection of electronic forms of the journals. After selecting the journals in electronic formats, the next step is to acquire that selected e-journals in the library.

2. MODES OF SUBSCRIPTION

⇒ E-Journals may be defined loosely as accessing full-text online journal articles. Full-text may be available through a variety of options: consortia based subscriptions of e-journals.
⇒ Subscriptions to online databases offered by vendors such as EbscoHost, ProQuest, InfoTrac etc.
⇒ Journal aggregators -- Electronic Journals Service (EJS) from Ebsco
⇒ Subscriptions from print/journal publishers directly -- El Sevier, Kluwer, Emerald, Wiley
⇒ Freely available online resources -- Directory of Open Access Journals (DOAJ: http://www.doaj.org/)
⇒ Pay-per-view access / document delivery -- Ingenta and individual publishers.

3. CURRENCY ISSUES: In purchasing items from abroad, following main issues arise:

1. Fluctuating exchange rates, which can upset the budget of a library.
2. Sudden rise in prices
3. Duties and taxes, such as customs duty. VAT (value-added tax) & sales tax.
4. Basis of calculating exchange rates:
   In India, this has been a serious problem. The best approach is to pay the amount using the rate of conversion, prevailing on the date of invoicing.

132

The vendor through a supplementary invoice should claim any subsequent increase in the rate of subscription at the then prevailing rates of conversion. In case there is reduction in the rate of subscription, the vendor should be accepted to issue a credit note, using the same rate of conversion as used in the original invoice.

4. ACQUISITION OF ELECTRONIC JOURNALS

During the last ten years or so, the process of acquisition of resources has undergone evolutionary changes. We are a witness to tremendous changes that have taken place. This due to many factors, such as exponential growth in the number of available resources, increasing variety, rising costs, different ways to acquire resources, declining purchasing power of special libraries; and changing environmental-social, economic, political and technological. Application of information technology has been the major factor responsible for our innovative approach in acquisition.

Acquisition of electronic journals or getting access to them needs proper planning and good management. The following consideration must be kept in view while subscribing the e-journal:

Technical Infrastructure: The existing technical infrastructure must be suitable. Some corporate networks have security systems, having firewalls that prevent access to some sites and downloading of certain files. This problem needs to be tackled before placing an order.

Licensing: The librarians must understand how the terms and conditions of licensing should have an impact on the use of journals.

Copyright: the librarian needs to understand as to how the local copyright restrictions correspond with the restrictions put down by the vendor.

Package of Products: Many electronic journals sold by the vendors as a package, consisting of an assortment of products. Hence the librarian may need only a specific product but he is forced to purchase all the products in the package, though there may be heavy discount as an incentive.

Vendor: It is found to be convenient, if e-books and e-journals are purchased from the same vendor. In that case, the same password can be used to access the products and services. The users can also access directly from the same web site of the vendor.

Trial Basis Run: It is always helpful to get the product on trial basis (not sample product) to identify and resolve problems, if any.

E-journals: Electronic journals (e-journals) are made available by publishers and vendors via web sites. The transient nature of e-journals is a matter of great concern to librarians. In acquiring e-journals, the librarian must carefully examine the following aspects:

Access to Back Issues (archiving): Will the back issues be accessible in future? What sort of guarantee does the vendor provide? Will there be additional charges for the same?

Content: Is there any difference in the content between electronic and print versions? If any, where lays the difference?

Restrictions: What are the restrictions on access? How many users are allowed access simultaneously? What are the restrictions on down loading? What are the restrictions on copying?

Web-based Reference Sources: Many reference resources (dictionaries, encyclopedias, directories, yearbooks, atlases, etc) are available on the web. This is often available free for a trial period and after the library has to pay ongoing charges. Often, the charges are similar to those of printed resources.

5. CATALOGUING

Many libraries are still pondering the appropriate way to catalogue electronic journals. The fact that many current titles are electronic versions of print titles has made it easier to incorporate these into the library's catalogue. This is done simply by adding a note to the existing cataloguing record informing of the existence of the digital copy and, if possible, by adding the hyperlink to permit direct access to the electronic copy. Although not a purely orthodox interpretation of the rules, it is an easy way around a sticky problem, similar to the way in which analogue reproductions have frequently been noted on the cataloguing record of the original in the past.

The problem, of course, is that where a facsimile on microfilm is identical in content and format to the original, the electronic version of a traditional journal is normally quite different from the original. One can assume that the differences will only increase as the digital version incorporates elements, which the print medium cannot accommodate: sound, moving images, interactive functionality, etc. It will only be a matter of time before the electronic version will require its own full cataloguing record to allow for a proper description of its unique properties.

While dual versions —print and electronic— are frequently described on the same cataloguing record, how are digital-only journals handled? The cataloguing of electronic journals challenges existing rules and procedures, since many of the basic elements of print journals are frequently missing, such as title pages, a fixed format, etc. There is little uniformity in the presentation of e-journals. For instance,

⇒ Some titles offer various options for viewing and printing the same text, for example ASCII, HTML, PDF, PostScript, etc.
⇒ Some journals publish articles as they are received, with the journal "issue" being compiled subsequently.
⇒ Some publishers' deposit published articles into a database, rather than maintaining them in discrete back issues.

These changes are so fundamental that the traditional definition of what is a serial is being reconsidered to distinguish between those that are issued in separate numbers (called "successively issued"), and those that are fed into a database ("integrating"). Most libraries concur that certain journals deserve full cataloguing in order to be able to give systematic and coherent access to them, integrated into the rest of the library's collection. While libraries may differ on exactly which titles warrant cataloguing, the typical selection includes:

1) Those that are subscribed to;
2) Those that are produced by the library or the institution it belongs to; and
3) Those that the library has an obligation to make available to the public (e.g., legal deposit material).

Other options chosen by some libraries are to list e-journals on the web page or to make them accessible in a database separate from the OPAC. In their recent article on how libraries handle free web resources, Estivill and Abadal discuss the present trends.

Given libraries' constant complaint about the high cost of serial publications, it is quite ironic that many free titles from major reputable entities are not being catalogued. While libraries cannot catalogue all free resources on the web, there are many out there that, were they issued on paper, would in fact be catalogued. But somehow, by virtue of being electronic, they are being bypassed.

134

6. STORAGE & PRESERVATION

Storing e-journals and the related issue of long-term preservation are topics currently under intense scrutiny. While some e-journals are in fact housed in the local institution, the overwhelming majority of titles are now being accessed via remote servers. Of course, some of the earliest experiences with e-journals used CD-ROMs for delivery, access and storage, such as *Adonis* http://www.adonis.nl/index.htm and *Ovid's Core Medical Collection* http://www.ovid.com/products/journals/cbc.cfm. Some publishers of online journals also provide CD-ROMs with the full year's contents, as a security copy. And numerous projects — such as *Elsevier's Tulip Project* and the EU-funded *Decomate project* http://www.cordis.lu/libraries/en/projects/decomate.html have explored models for local storage and access. In fact, the Decomate system is used by the Universitat Autònoma de Barcelona for providing access to journal literature stored in local servers http://decomate.uab.es. Although systems for locally stored information give the library greater control over the information and its accessibility, it does so at a high cost to the institution in terms of development, maintenance and storage.

As Internet connections have become more powerful and dependable, the current model of remote access has come to be accepted as the standard model for distribution and access of e-journal contents. As such, the library never physically obtains the contents, which remain at the supplier's site for direct access by the user. This has profound implications for long-term preservation, as well as for usage. A library's paper copy of a journal remains in its possession regardless of whether the subscription has been cancelled or if the title is no longer being published. The shift from local to supplier-based storage removes the assurance of permanent access. What will happen to the contents of the journals —licensed rather than purchased— if the library decides to cancel the subscription, if the publisher drops the title or if the publisher ceases to exist? Publishers are in no position to take on the challenge of permanent storage and so models are being sought.

Yale University's library and Elsevier Science recently announced a project for developing a digital archive for the 1,100 journals published electronically by Elsevier Science http://www.elsevier.nl/inca/esav/shownews, OCLC's Electronic Collections Online (ECO) service http://www.oclc.org/oclc/eco/archive.htm guarantees libraries permanent access to contents that had been acquired, as well as the future migration of these contents to new platforms, formats, etc. It does this by requiring that the publisher deposit the collections with OCLC and charging libraries a separate maintenance fee apart from the subscription, which must be paid regardless of whether the title is still alive, or being subscribed to. Of course, as a library consortium, OCLC is in a better position to commit to long-term storage since its members decide on the organization's goals and strategy based on their own interests and needs. This is quite different from most other service providers who are market- and profit-driven.

Assuring that users have access to the contents is not the only issue. Of greater concern is how to guarantee the preservation and usability of the information, threatened by factors that are both inherent in the storage media, as well as external to it, such as technological advances. Generational changes leave documents created within a certain technological environment —software, hardware, operating system- unreadable as the technology evolves. A document left behind —for instance an article prepared using the program WordStar and stored on a 5 1/2 inch diskette in 1988— is considered an "orphan" if subsequent technological changes have obliterated the components necessary for it to be read. The suggested measures for ensuring permanent access to digital content include refreshing data by passing it from one support to another (to avoid degeneration of the physical medium); continual migration of data from one platform to another in order to keep up with both technological changes and software evolution, and prevent it being "orphaned"; emulation, by which systems of the future will be able to interpret data created in the past, by replicating older platforms in order to preserve the "look and feel" of original documents; and computer museums in which entire systems will be maintained in order to read documents

Bearing in mind the speed at which systems change —calculated at 3 years for hardware and 18 months for software— it is obvious that much serious attention need be given to this problem in order to assure that today's e-journals will be readable long into the future —hopefully for as long as 350 years— in the same way that readers can still consult the earliest journals from the 17th century.

Obviously long-term storage and preservation of digital materials go beyond e-journals and affect all digitally created material. The purely technical questions are frequently surpassed in complexity by other issues, such as metadata, data integrity and authenticity, copyright, standards, etc., all of which are beyond the reach of this article. There are a number of interesting international initiatives exploring all aspects of the issue, including the CEDARS project in the UK http://www.leeds.ac.uk/cedars & the EU-funded NEDLIB project http://www.kb.nl/coop/nedlib.

7. ACCESSING E-JOURNALS

The user faced problems during the e-journal access in library. It is because of several factors. It may be technical (Internet congestion, bandwidth disruption, utility software unavailability etc.), psychological, personal and the institutional (e.g. library staff support). The subject of use and usability spans several issues, from the types of usage permitted by the license agreement to how libraries make e-journals available to their users and how the users react to the transition from print to Web. In the following description, we will be able to reach on the factors or reasons which are chiefly responsible. A study at Lucknow University of the libraries subscribing e-journals reveals the following basic factors, which adhere before the scientists, research scholars and other information seekers during the access of e-journals.

Users in the library including Scientists, Researcher Scholars and the staff of the laboratories encountered following problems during the access of e-journals.

Technical Problems;
⇒ *Slow speed of Internet*: sometime Internet speed become slowly and the time taken in a site opening irritates to the users. Speed also depends on the bandwidth of Internet.
⇒ *Disconnection*: because of congestion in network of ISP it disconnected several times and hanged too.
⇒ *Sometime access* to the full text of journals is available from the publisher's website, but it requires validation by IP address and is therefore only accessible on nominated PCs.
⇒ *Unavailability of utility software*. E.g. acrobat reader is essential to read the downloaded Pdf files.

Psychological Problems:
⇒ Attitudinal problems: generally new user feel shy during the Internet use and they don't ask to the librarian about the problems encountered during the Internet and e-journal access. For example if once access code is asked by the online publisher and in first attempt they did not get it well from the library staff, then generally don't ask it again.
⇒ *Misconception about Internet*: still users have misconception about the Internet and its facilities.

Overflow of Information:
Over the Internet lots of information and information sources are available and there is huge information garbage. There is no option to select the right and adequate information. Its mean to say that information is available over Internet very frequently and reconviction of retrieved information is not an easy task.

Problems from Library Staff:
⇒ *Untrained Staff:* Sometimes because of untrained library staff, they have to spend their precious time over some unnecessary work and procedures.
⇒ *Not helpful:* the attitude of library staff also creates the problem for the e-journal access as if user is accessing the e-journal and time for closing for the library is near then he would usually insist to stop the accessing the e resources.

Use of online accessories/tools:
⇒ *Familiarity with Information Gateways:* Its is found that very often scientist doesn't know about the information gateways and the internet indexes etc. as majority of users believe that these gateways are very helpful and they save lots of time of users. But they don't have much known information gateways. They prefer to use only those, which were, suggested them by their seniors.
⇒ *Search Engines:* information user generally used only the goggle and msn search engines which are automatically appears on putting the search expression in the dialog box. It is found that they don't know about the types of search engines yet they believe that search engines are helpful in searching the urls of the seeking organization.

Webliographic control:
There is no webliographic control possible that means there is no record that how much documents are online on any topic. It is almost impossible for any professional to be well versed about the total available online information resources. Information explosion has been encountered such an extent that today experts and professors possess mere a bit of knowledge.

CONCLUSION
Here in this paper, I have tried to give a glimpse of the major aspects related to e-journal and its merits and demerits in a library. We have now evolved to a point at which many scholarly journals are available in electronic form on the web, greatly extending the availability of these resources over time and distance. But it remains to be seen if the success of this new delivery means will ultimately fatally undermine the status of the traditional journal format, as we have come to know it over 350 years.

Today libraries make more of their services and collections available via web, a logical consequence has been that their attendance levels have declined in the libraries. Professors, researchers and students alike can access the needed information from home or office. While overall this is added convenience to users has been seen as a positive result, it has also produced the phenomenon of "the hidden user".

Libraries have devoted a great amount of effort over the past decade to aligning their services to the users' true needs. Now that users are disappearing from the library, it becomes all the more difficult to know exactly who they are, to understand their information needs, to aid in information seeking, and to evaluate whether they are obtaining appropriate results. Hence at the end, hopefully the library and information professional will be benefited by the above description about e-journal, its various features and problem encountered in its subscription and use.

References:
1. Franks, John. What is an Electronic Journal? Department of Math's, Northwestern University, Evanston, IL. January 1993.
2. Keyhani, Andrea. 'The Online Journal of Current Clinical Trials: An Innovation in Electronic Journal Publishing Database, 16(1): 14-23 February 1993.
3. Munsi, usha mujoo, visualizing an Electronic Library/ Usha Mujoo-Munshi. IN Vistas in Library and Information Systems and Networks: a festschrift to professor S. M.

5. Paul pedley; Electronic Journals 'Managing Information' 2001
6. Rathie, Susan. 'Electronic Journals and Peer Review: Perils and Promises'. University of British Columbia. 1994
7. http://www.library.uiuc.edu/icsu/hall.htm
8. http://www.ub.es/biblio/bid/06keefe2.htm
9. http://is.unn.ac.uk/lmri/CurrentResearch/JUBILEE/jubilee2.htm

By : *Pragya Pandey ,librarian,City Montessori College,Aliganj,Lucknow E-mail:praggyapandey@yahoo.cor
** Kamal Chaurasia, Librarian at Moti Lal Rastogi School of Mgmt. Lucknow-8. Email:
klib_522@yahoo.com

138

NASSDOC/GIDS TRAINING PROGRAME
ON
INFORMATION RESSORUCE
MANAGEMENT FOR SOCIAL SCIENTISTS
IN EMERGING KNOWLEDGE ERA
13-15 FEBRUARY, 2007

LISTOF PARTICIPANTS

SN	Participants	Phone/Fax/e-mail
1.	Ms. Tulsi Mehta Lecturer, Career Convent GIDS, Lucknow	2327783
2.	Mr. Yogendra Bechain Teacher Dayanand P.G. College Orai, Jalaun-285001	9936715509 madhukar26@rediffmail.com
3.	Ms. Shruti Pandey Librarian SHEPA. Saraswati Vihar, Nibia, Bachhaow Varanasi	9450540769 shruti shepa@rediffmail.com
4.	Ms. Poonam Deo Library & Information Assistant Dental Library K.G.U.D.S. King George University of Dental Science Lucknow	9839107116 deep-aradhana@yahoo.co.in
5.	Mr. Sushil Kumar Tripathi Librarian B.R.D.S.V.M. Sr. Sec. School Sect.-Q, Aliganj Lucknow	9450466099
6.	Mr. Kamal Kumar Chaurasia Librarian Moti Lal Rastogi School of Management D-23, UPSIDC, Industrial Area Sarojini Nagar Lucknow	(0522) 2437174

139

7.	Mr. Yogendra Mishra NBRI Rana Pratap Marg Lucknow	(0522) 2205836 9450357985 yshiksha@yahoo.co.in
8.	Ms. Suman Jaiswal Student B.I.T. Akansha Girls Hostel Raguveer Colony Mahmod ganj Varanasi	9839858599
9.	Ms. Deepika Srivastava Student B.I.T. Varanasi C/O Sri Rakesh Srivastava Tarakpur Mirzapur	
10.	Mr. Vinod Kumar Mishra Librarian Govt. Polytechnic Faizabad Road Lucknow	0522-2700680 9415087848
11.	Mr. Rakesh Kumar Srivastava Assitt. Librarian Era's Medical College & Hospital Hardoi Road Lucknow	Fax-91-0522-2247824 0522-2248122 0522-2248123 rakeshlucknow51@rediffmail.com
12.	Ms. Pragya Pandey Librarian CMS, Sect-O, Aliganj Lucknow	9839838184 pragypandey@yahoo.com
13.	Mr. Pankaj Kumar Lecturer SAMA, Lucknow Bakshi Ka Talab Navikot Nandana, Sitapur Road Lucknow	0522-2333992 9839137506
14.	Mr. Sanjeev Mishra Executive Library SAMA, Lucknow Bakshi Ka Talab, Navikot Nandana, Sitapur Road, Lucknow	0522-2333992 9936045594 san-lko2002@yahoo.com

140

15.	Dr. Raghu Ram Sr. Lecturer Ambedkar Central University Lucknow	Fax-0522-2440821 9415023195 raghup2006@yahoo.co.in
16.	Ms. Ratna Singh Researcher C/o Dr. Amar Singh 537 Cha/15, Sherwani Nagar Sitapur Road Lucknow	0522-2733231 niti-ddun@yahoo.co.in
17.	Ms. Amita Mishra Asstt. Librarian Sherwood College of Management Sect-25, Indira Nagar Lucknow	4. 9919744433 ami- amita16484@rediffmail.com
18.	Kunwar Abhishek Pratap Singh Library Assistant U.P. Gandhi Smarak Nidhi Gandhi Bhawan Lucknow-01	0522-2625396 (O) kunwar432@yahoo.co.in
19.	Mr. Sushil Kumar Srivastava C-44, Raj Bhawan Compound Hazratganj Lucknow	9450638732
20.	Mr. Dharmendra Singh Library Assistant IIM Lucknow	Fax-0522-2734025-26 0522-2736963 9415432583
21.	Mr. Sharad Kumar Sonker Lecturer Baba Saheb Bhimrao Ambedkar University Vidya Vihar, Rae Bareilly Road Lucknow	9838616581 sksonkar@yahoo.co.in
22.	Mr. Amitabh Saran Assistant Librarian (Judges Library) High Court Bench Lucknow	Fax-0522-2272328 0522-2627395 ext. 256 9415400063 Sarn-amitabh@yahoo.com
23.	Mr.Ateeq-ur-Rahman Assistant Librarian Central Library Integral University, Kursi Road Lucknow	0255- 2890809 9919766369 Ateeqrahmanlib@yahoo.co.in

24.	Ms. Gayatri Singh Cataloguer I.E.T., Sitapur Road Lucknow	
25.	Mr. Sanjay Kumar Singh Sr. Lecturer I.E.T., Sitapur Road Lucknow	Fax-0522-2361631 945136411 sanjaiasdiet@rediffmail.com
26.	Ms. Sunita Pandey Assistant Librarian Ram Swaroop Engineering College Lucknow	0522-2768966 9918524068
27.	Ms. Sharmila Ghosh Consultant, Human Rights Law Network, New Delhi B-606,CDOT Apartments GH-2, Sect-56 Gurgaon (Haryana)	0124-2574553 9818742924 milaghosh@gmail.com mana-ghosh2002@yahoo.co.uk
28.	Dr. Seema Shukla Assistant Librarian Centre for Social Studies South Gujrat University Campus Udhana-Magdalla Road Surat-395007 (Gujrat)	0261-2227173/74 shuklaseema2@rediffmail.com
29.	Dr. B.P. Srivastava Librarian D.S.N. College Unnao	
30.	Ms. Indira Madan Librarian Doordarshan Kendra 24, Ashok Marg Lucknow	0522-2288425 imdd.lucknw@yahoo.co.in

142

28.	Ms. Ila Joshi Librarian NIPCED	0522-2362543 0522-2362643
31.	Ms. Poonam Mishra Librarian ICMRT Lucknow	0522-2716431
32.	Mr. Pradeep Singh Student, M.L.Sc. L.U. Lucknow	0522-2332177
33.	Ms. Purnima Pandey Librarian Lucknow Public School 21/304, Gandhar, Sahara State, Jankipuram, Lucknow	9235702679 Purnima1467@yahoo.co.in
34.	Ms. Anjali Gulati Lecturer Dept. of Lib. & Inf. Sci. I.T. College Lucknow	0522-2662214

LIST OF RESEARCH SCHOLAR OF GIDS

1. Ms. Pratima Ghosh
   Research Scholar
   GIDS
   Lucknow

2. Ms. Atika Khatoon
   Field Investigator
   GIDS
   Lucknow

3. Ms. Pratibha Raj
   ICSSR Fellow
   GIDS
   Lucknow

143